

SENASTIKA 2024

SEMINAR NASIONAL TEKNOLOGI INFORMASI
DAN TEKNIK INFORMATIKA



GOR ACC CUNDA
UNIVERSITAS MALIKUSSALEH
LHOKSEUMAWE

17 OKTOBER 2024

[Current](#) [Archives](#) [About](#) ▾

[Home](#) / [Archives](#) / Vol. 1 No. 1 (2024): Seminar Nasional Teknologi Informasi dan Teknik Informatika (SENASTIKA) 2024

Vol. 1 No. 1 (2024): Seminar Nasional Teknologi Informasi dan Teknik Informatika (SENASTIKA) 2024

Published: 2024-11-01

Articles

ANALISIS DAN PERBANDINGAN METODE CNN DAN SVM DALAM MENDETEKSI BATIK NUSANTARA

Nanda Nan Arif, Muslimatul Magfirah, Fuzna Febriani, Rodhatul Jannah, Munirul Ula

[PDF \(Bahasa Indonesia\)](#)

DIAGNOSIS PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

David Fadlianda, Aditya Prananto, Cut Anggel Eriska, Syifa Anjanira, Nada Syadzwin, Munirul Ula

[PDF \(Bahasa Indonesia\)](#)

IMPLEMENTASI DETEKSI BERITA CLICKBAIT BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR

SENASTIKA Universitas Malikussaleh

ANALISIS DAN PERBANDINGAN METODE CNN DAN SVM DALAM MENDETEKSI BATIK NUSANTARA

Nanda Nan Arif¹, Muslimatul Magfirah², Fuzna Febriani³, Rodhatul Jannah⁴, Munirul Ula⁵

Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: ¹nanda.210170138@mhs.unimal.ac.id, ²muslimatul.210170141@mhs.unimal.ac.id,

³fuzna.210170122@mhs.unimal.ac.id, ⁴rodhatul.210170145@mhs.unimal.ac.id, ⁵munirulula@mhs.ac.id

Abstrak

Batik, sebagai warisan budaya Indonesia, memiliki kekayaan motif yang unik di setiap daerah. Penelitian ini bertujuan untuk mengklasifikasikan secara otomatis motif batik dari berbagai daerah seperti Aceh, Bali, Betawi, Dayak, Papua, dan lainnya, masing-masing dengan ciri khas yang unik, menggunakan metode Convolutional Neural Network (CNN) dan Support Vector Machine (SVM). Dengan membandingkan kinerja kedua metode, penelitian ini bertujuan untuk menemukan pendekatan yang paling efektif dalam mengidentifikasi karakteristik visual yang membedakan jenis batik. Hasil penelitian menunjukkan bahwa metode SVM memberikan kinerja yang lebih baik dibandingkan CNN dalam skenario dengan hanya 20 kategori batik dan masing-masing kategori memiliki 100 gambar. SVM terbukti lebih efisien dalam mengenali pola-pola visual dengan jumlah data yang terbatas, sementara CNN memerlukan lebih banyak data untuk mencapai hasil yang optimal. Temuan ini berpotensi diaplikasikan dalam autentikasi batik, perlindungan kekayaan intelektual, serta promosi budaya batik secara digital. Penelitian ini mengevaluasi efektivitas kedua metode dalam mengenali pola visual batik, dan menunjukkan bahwa SVM lebih unggul dalam situasi dengan dataset yang lebih terbatas

Kata Kunci : *Convolutional Neural Network, Support Vector Machine, Batik, Klasifikasi, Corak*

Abstract

Batik, as part of Indonesia's cultural heritage, possesses unique patterns specific to each region. This study aims to automatically classify batik motifs from various regions such as Aceh, Bali, Betawi, Dayak, Papua, and others, each with its distinctive characteristics, using Convolutional Neural Network (CNN) and Support Vector Machine (SVM) methods. By comparing the performance of these two methods, this research seeks to identify the most effective approach for distinguishing the visual characteristics of different batik types. The results of the study show that the SVM method outperforms CNN in a scenario with only 20 batik categories and 100 images per category. SVM proved to be more efficient in recognizing visual patterns with a limited amount of data, while CNN requires more data to achieve optimal results. These findings have potential applications in batik authentication, intellectual property protection, and the digital promotion of batik culture. This research evaluates the effectiveness of both methods in recognizing batik visual patterns and demonstrates that SVM performs better in situations with a more limited dataset.

Keywords : *Convolutional Neural Network, Support Vector Machine, Batik, Classification, Pattern*

1. PENDAHULUAN

Bangsa Indonesia merupakan negara yang terdiri dari banyak suku bangsa yang tersebar dari Sabang sampai Merauke, dimana setiap suku bangsanya mewariskan hasil kesenian tradisional yang berbeda beda. Keragaman motif batik ini menjadikannya aset budaya yang tak ternilai. Salah satunya adalah batik, Batik merupakan salah satu pakaian tradisional asli Indonesia yang telah dikenal secara luas di wilayah Indonesia dan telah ditetapkan oleh badan dunia PBB yakni UNESCO sebagai warisan budaya bangsa Indonesia. Pengakuan batik oleh UNESCO sebagai warisan budaya memuat amanah untuk memelihara batik sebagai warisan budaya tak benda dari Indonesia yang dikenal dengan Intangible Cultural Heritage [1]. Batik juga salah satu pakaian yang di ekspor ke negara-negara luar. Batik merupakan kain bergambar yang pembuatannya secara khusus dengan menuliskan atau menerapkan lilin malam pada kain, kemudian pengolahannya diproses dengan cara tertentu yang memiliki kekhasan tersendiri di setiap daerahnya. Proses pembuatan batik ini sangat lama, memakan waktu paling cepat dua minggu jika cuaca bagus.

Indonesia memiliki berbagai jenis batik dari berbagai daerah, seperti Batik Aceh, Bali, Betawi, Dayak, dan Papua, dan lain-lain. Masing-masing batik tersebut memiliki ciri khas unik yang mencerminkan kekayaan budaya dan sejarah lokal. Batik Aceh cenderung menggunakan motif geometris dan flora dengan sentuhan budaya Islam, serta warna-warna cerah seperti merah, hijau, dan kuning, yang memberikan kesan glamor. Beberapa motif khasnya meliputi motif tulak angen, Pinto Aceh, dan pucok reubong [2]. Sementara itu, Batik Bali memiliki motif yang lebih rumit dengan pengaruh agama Hindu dan mitologi lokal, seperti kisah para dewa. Motifnya sering kali memadukan ornamen khas Bali seperti naga, rusa, dan bangau, dengan unsur flora dari luar Bali [3]. Kemudian, Batik Dayak berasal umumnya memiliki motif yang terinspirasi dari alam, seperti motif tumbuhan, hewan, serta

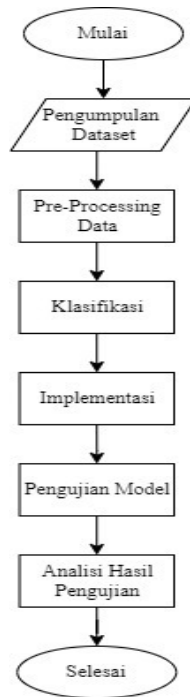
simbol-simbol adat Dayak, mencerminkan hubungan yang kuat dengan alam dan spiritualitas. Batik Betawi sering menampilkan motif ondel-ondel, monas, dan flora khas Betawi, menunjukkan kearifan lokal serta semangat kebhinekaan. Sementara Batik Papua cenderung memiliki corak yang lebih berani dan tegas, dengan motif khas seperti burung cendrawasih, patung asmat, serta simbol-simbol budaya Papua yang kaya akan tradisi. Perbedaan motif batik menjadikannya menarik untuk diteliti dalam pengenalan dan klasifikasi citra. Dengan perkembangan teknologi, otomatisasi identifikasi batik berdasarkan motif visualnya semakin penting untuk pelestarian dan promosi budaya secara digital. Meskipun batik Indonesia memiliki nilai budaya yang tinggi dan keindahan khas banyak orang, baik di dalam maupun luar negeri, belum sepenuhnya mengenal berbagai jenis batik. Hal ini disebabkan oleh kurangnya edukasi dan promosi, serta variasi motif batik yang sering membuat masyarakat salah mengidentifikasi asalnya. Permasalahan lain seputar batik adalah adanya klaim dari negara lain. Hal ini berarti membuat batik Indonesia butuh ekstra perlindungan dari berbagai upaya klaim, pencurian motif, maraknya printing batik dan minimnya akses dan distribusi juga menghambat pengenalan batik dari beberapa daerah, meskipun Indonesia memiliki banyak penghasil batik. Salah satu pendekatan yang dapat diterapkan dalam mendeteksi motif kain batik adalah melalui pengolahan citra digital yang didukung oleh metode pembelajaran mesin. *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)* adalah dua metode yang sering digunakan dalam pengolahan citra dan klasifikasi pola.

Tujuan utama dari penelitian ini adalah mengembangkan sebuah metode otomatis yang dapat mengidentifikasi dan mengklasifikasikan secara akurat berbagai macam motif batik Indonesia berdasarkan citra digital. Dengan adanya metode ini, diharapkan dapat memberikan kontribusi signifikan dalam upaya pelestarian batik melalui autentifikasi dan identifikasi yang lebih mudah. Selain itu, metode ini juga dapat dimanfaatkan untuk memperluas jangkauan promosi batik secara digital, sehingga dapat menjangkau pasar yang lebih luas dan menarik minat generasi muda terhadap warisan budaya Indonesia.

Convolutional Neural Network (CNN) merupakan sebuah algoritma yang terdiri dari proses ekstraksi ciri menggunakan konvolusi dan proses klasifikasi yang menggunakan neural network. CNN juga merupakan bagian dari metode Deep Learning yang dapat digunakan untuk mendeteksi dan mengenali sebuah objek pada sebuah citra digital. Pada dasarnya *Deep Learning* merupakan bagian dari *Machine Learning* yang menerapkan ide konsep jaringan syaraf manusia yang disebut neural network dengan lapisan yang tersmbunyi yang lebih banyak [4]. CNN mampu mengenali pola kompleks, seperti tekstur dan bentuk yang terdapat pada citra kain batik. Keunggulan CNN adalah kemampuannya melakukan feature learning secara bertingkat, sehingga memberikan hasil yang baik dalam tugas klasifikasi citra. *Support Vector Machine (SVM)* adalah algoritma pembelajaran mesin yang berfungsi memisahkan data ke dalam dua atau lebih kelas berdasarkan margin terbesar. SVM dikenal efektif dalam mengatasi masalah klasifikasi, terutama ketika data yang tersedia terbatas atau memiliki dimensi tinggi. SVM telah berhasil digunakan untuk klasifikasi motif batik. SVM memiliki popularitas besar dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, dan lebih efektif efektif [5]. SVM merupakan metode untuk klasifikasi yang mempunyai kemampuan generalisasi lebih tinggi dan dapat memberikan performansi yang lebih baik bila dibandingkan metode-metode pengklasifikasian lainnya. Penelitian ini diharapkan berkontribusi signifikan secara teoritis dan praktis. Secara teoritis, memperkaya pemahaman dalam pengolahan citra dan *deep learning* untuk klasifikasi motif batik. Metode yang dikembangkan dapat menjadi alternatif lebih akurat dalam mengklasifikasikan citra dengan variasi tinggi. Secara praktis, mendukung pelestarian batik melalui identifikasi dan autentifikasi motif yang efisien, serta mempromosikan batik secara digital untuk menjangkau pasar lebih luas dan menarik minat generasi muda. singkatkan lagi

2. METODE PENELITIAN

Dalam melakukan klasifikasi pada motif Batik, ada beberapa tahapan yang dilakukan dalam penelitian ini yang mana ditampilkan pada Gambar1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Dataset

Ini membandingkan kinerja *Convolutional Neural Network* (CNN) dan *Support Vector Machine* (SVM) dalam mengklasifikasi citra kain batik Aceh dan Bali. Dataset yang digunakan terdiri dari 2000 citra, meliputi 100 jenis batik Aceh, 100 jenis batik Bali Barong, 100 jenis Batik Betawi, 100 jenis Batik DKI, dan 100 jenis Batik Papua Tifa dan masih banyak lagi. Masing-masing jenis batik direpresentasikan oleh beberapa citra untuk melatih dan menguji model.

Tabel 1. Detail Dataset

Sampel	Jumlah Data	Sumber Data
Batik Aceh	100	Kaggle
Batik Bali Barong	100	Kaggle
Batik Bali Merak	100	Kaggle
Batik DKI	100	Kaggle
Batik Jawa Barat	100	Kaggle
Batik Jawa Timur	100	Kaggle
Batik Kalimantan	100	Kaggle
Batik Lampung	100	Kaggle
Batik Madura	100	Kaggle
Batik Maluku	100	Kaggle
Batik NTB	100	Kaggle
Batik Papua Asmat	100	Kaggle
Batik Papua Cendrawasih	100	Kaggle
Batik Papua Tifa	100	Kaggle
Batik Solo	100	Kaggle
Batik Sumatera Barat	100	Kaggle
Batik Sumatera Utara	100	Kaggle
Batik Yogyakarta Parang	100	Kaggle
Batik Yogyakarta Kawu	100	Kaggle
Total	2000	

Berikut merupakan menunjukkan detail motif batik dari beberapa wilayah di Indonesia.



Gambar 2. Motif Batik Nusantara

2.2 Pre – Processing

Menurut buku "*Introduction to Data Mining*" karya Pang-Ning Tan, Michael Steinbach, dan Vipin Kumar (2005), Pre-processing data adalah proses yang melibatkan transformasi data mentah ke dalam format yang lebih cocok untuk analisis dan pemodelan. Tujuan dilakukannya pre processing ini agar bisa memperbaiki atau meningkatkan kualitas citra dan menyesuaikan citra supaya dapat meningkatkan performansi dari sistem.

2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai salah satu metode machine learning yang bekerja dengan prinsip *Structural Risk Minimization* (SRM). Prinsip SRM bertujuan untuk menemukan hyperplane terbaik yang memisahkan dua buah kelas pada *input space*. Metode ini menggunakan hipotesis berupa fungsi linier dalam ruang fitur berdimensi tinggi, dengan mengimplementasikan proses *learning* yang berdasarkan pada teori pembelajaran statistik [6]. Support Vector Machine (SVM) adalah teknik yang digunakan untuk prediksi, baik dalam kasus klasifikasi maupun regresi. Prinsip dasar SVM adalah *linear classifier*, yaitu untuk kasus klasifikasi yang secara linier dapat dipisahkan. Namun, SVM telah dikembangkan agar dapat menangani masalah non-linier dengan memasukkan konsep *kernel* pada ruang berdimensi tinggi. Dalam ruang berdimensi tinggi ini, SVM mencari hyperplane yang memaksimalkan jarak (margin) antara kelas data yang berbeda.

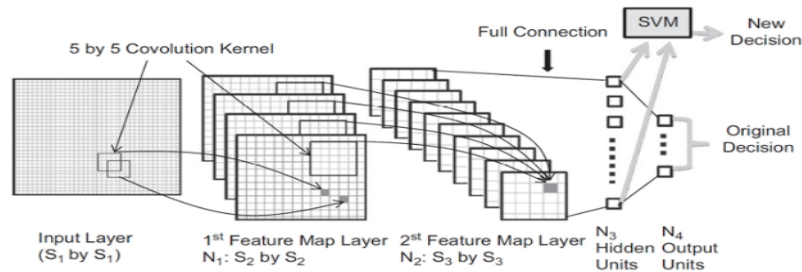
2.4 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah salah satu jenis neural network yang biasa digunakan pada data gambar yang digunakan untuk mendeteksi dan mengenali objek. *Convolutional Neural Network* (CNN) dirancang untuk melakukan pekerjaan dalam memproses data ke dalam bentuk array, misalnya gambar berwarna yang terdiri dari tiga 2D array yang memiliki intensitas piksel di tiga saluran warna, Convolutional Neural Network memanfaatkan sifat nature signals yaitu koneksi lokal, share wheights, pooling, dan penggunaan banyak lapisan. Terdapat empat model dalam proses pengolahan citra, yaitu Convolution Layer, Pooling Layer, Dropout Layer, Fully Connected Layer [7]. Pada tahap pemrosesan citra, citra akan melewati beberapa langkah, dimulai dengan penerapan filter. Citra kemudian direduksi dengan cara mengambil nilai terbesar dari setiap grid (max pooling) untuk mengurangi dimensi data. Selanjutnya, dilakukan pencegahan overfitting melalui teknik tertentu. Setelah itu, data ditransformasikan ke dimensi yang lebih sesuai untuk proses klasifikasi. Agar hasil klasifikasi akurat, model ini memerlukan data pelatihan yang berkualitas. Salah satu komponen penting dalam Convolutional Neural Network (CNN) adalah Convolutional Layer. Pada lapisan konvolusi ini, filter digunakan untuk mendeteksi karakteristik dari objek atau citra, menghasilkan representasi linier dari citra input yang relevan [8]. Ada beberapa parameter penting dalam setiap lapisan, seperti filter, stride, dan padding. Stride mengontrol bagaimana filter bergerak melintasi data input, dengan setiap pergerakan ditentukan berdasarkan ukuran piksel yang diinginkan. Padding menambahkan piksel di sekitar data input dengan nilai tertentu agar ukuran output tidak terlalu kecil, sehingga informasi penting tidak hilang selama proses konvolusi.

2.5 Klasifikasi

CNN-SVM adalah model gabungan dari CNN dan SVM dengan mengganti output layer (lapisan keluaran terakhir) dari model CNN dengan pengklasifikasi SVM. Pada output layer model CNN, merupakan lapisan yang melakukan perkiraan probabilitas untuk sampel masukan. Pada Proses CNN-SVM, dengan gambar input yang

dinormalisasi dan dipusatkan lalu dikirim ke lapisan input, model CNN dengan lapisan keluaran diatur beberapa epochs sampai proses pelatihan menyatu. Kemudian SVM dengan kernel yang digunakan pada penelitian ini yaitu kernel linear dan kernel RBF (Radial Basis Function) akan menggantikan lapisan keluaran. SVM mengambil keluaran dari hidden layer (lapisan tersembunyi) sebagai fitur vector baru untuk pelatihan. Setelah pengklasifikasi SVM telah terlatih dengan baik, maka dilakukan tugas pengenalan dan membuat keputusan baru pada gambar pengujian dengan fitur yang diekstraksi secara otomatis [9]. Gabungan model CNN-SVM dapat dilihat pada gambar dibawah ini.



Gambar 3. Hyperplane Support Vector Machine

2.6 Bahasa Pemrograman

Python merupakan bahasa pemrograman yang terkenal yang bisa dipakai untuk beberapa hal, dimulai dari membuat website, mengkaji data, scripting, sampai pembuatan game. Bahasa pemrograman python dirancang dengan berbagai macam multifungsi. Python juga mempunyai kode yang lengkap, dan mudah dipahami [10].

3. HASIL DAN PEMBAHASAN

Hasil penelitian dengan google colab untuk mengembangkan model Convolutional Neural Network (CNN) dan Support Vector Machine (SVM) ini menggunakan beberapa library diantaranya yaitu: Tensorflow, Scikit, Seaborn, Zipfile, Shutill, Matplotlib, Numpy, Keras processing, dan Cv2. Library ini menyediakan berbagai alat dan framework untuk membuat, melatih, dan menjalankan model machine learning dan neural network. Selain itu, dengan integrasi Keras sebagai high-level API, Tensorflow mempermudah pengguna dalam memulai dan mengembangkan aplikasi deep learning dari konseptualisasi hingga deployment [11]. Library Matplotlib berfungsi untuk visualisasi data. Matplotlib digunakan untuk membuat berbagai jenis plot seperti line plot, scatter plot, bar plot, histogram, dan lain sebagainya. Library ini memberikan kontrol yang luas terhadap elemen-elemen dalam gambar yang dihasilkan, seperti garis, teks, label sumbu, dan lainnya, sehingga memungkinkan pengguna untuk membuat visualisasi yang sangat disesuaikan. scikit (Scikit-Learn, yang sering disebut hanya sebagai Scikit, adalah pustaka (library) open-source berbasis Python yang digunakan untuk machine learning dan analisis data. Scikit-Learn memudahkan implementasi berbagai algoritma machine learning yang umum, seperti klasifikasi, regresi, clustering, dan pengurangan dimensi, serta menyediakan berbagai alat untuk pemrosesan data dan evaluasi model [12].

NumPy (Numerical Python) adalah pustaka (library) fundamental dalam ekosistem Python yang digunakan untuk komputasi numerik. NumPy menyediakan dukungan untuk array multidimensi, serta berbagai fungsi matematika yang efisien untuk operasi aljabar linear, statistik, dan transformasi data. Keras adalah pustaka open-source yang dibangun di atas TensorFlow, digunakan untuk membangun dan melatih model deep learning. Salah satu komponen penting dalam Keras adalah modul Keras Processing, yang menyediakan alat untuk pra-pemrosesan data sebelum data tersebut digunakan untuk melatih model. Pra-pemrosesan ini sangat penting karena kualitas data yang digunakan untuk melatih model sangat mempengaruhi performa model tersebut [13]. Dan yang terakhir Cv2, merupakan modul dalam pustaka OpenCV (Open Source Computer Vision Library), yang merupakan pustaka open-source untuk aplikasi computer vision dan pemrosesan gambar. Pustaka ini digunakan untuk berbagai tugas seperti deteksi objek, pengenalan wajah, segmentasi gambar, pelacakan gerak, serta pemrosesan video [14].

Adapun pra-pemrosesan data dalam penelitian ini melibatkan persiapan 2000 data gambar motif batik Nusantara yang dibagi menjadi 20 daerah, dengan masing-masing daerah memiliki 100 gambar motif batik. Enam puluh persen dari keseluruhan data digunakan untuk pelatihan, dua puluh persen digunakan untuk data validasi dan dua puluh persen lagi digunakan untuk pengujian.

Tabel 2. Pembagian dataset

Dataset	Data Trining 60%	Data Validasi 20%	Data Testing 20%
Batik Aceh	60	20	20
Batik Bali Barong	60	20	20
Batik Bali Merak	60	20	20
Batik DKI	60	20	20
Batik Jawa Barat	60	20	20
Batik Jawa Timur	60	20	20
Batik Kalimantan	60	20	20
Batik Lampung	60	20	20
Batik Madura	60	20	20
Batik Maluku	60	20	20
Batik NTB	60	20	20
Batik Papua Asmat	60	20	20
Batik Papua Cendrawasih	60	20	20
Batik Papua Tifa	60	20	20
Batik Solo	60	20	20
Batik Sumatera Barat	60	20	20
Batik Sumatera Utara	60	20	20
Batik Yogyakarta Parang	60	20	20
Batik Yogyakarta Kawu	60	20	20
Total	1200	400	400

```
[7] # Assign directories for training and validation
train_dir = os.path.join('batik_split', 'train')
val_dir = os.path.join('batik_split', 'val')

# Load data train dan validation
train_images, train_labels = load_data(train_dir)
val_images, val_labels = load_data(val_dir)

# Mengubah gambar menjadi vektor 1D
train_images = train_images.reshape(len(train_images), -1)
val_images = val_images.reshape(len(val_images), -1)

# Encode label
label_encoder = LabelEncoder()
train_labels_encoded = label_encoder.fit_transform(train_labels)
val_labels_encoded = label_encoder.transform(val_labels)

# Membuat dan melatih model SVM
svm_model = SVC(kernel='linear', C=1.0, random_state=42)
svm_model.fit(train_images, train_labels_encoded)

# Melakukan prediksi
val_predictions = svm_model.predict(val_images)

# Menghitung akurasi
accuracy = accuracy_score(val_labels_encoded, val_predictions)
print("Akurasi SVM pada data validasi:", accuracy)

Akurasi SVM pada data validasi: 0.1975
```

Gambar 4. Hasil Support Vector Machine (SVM)

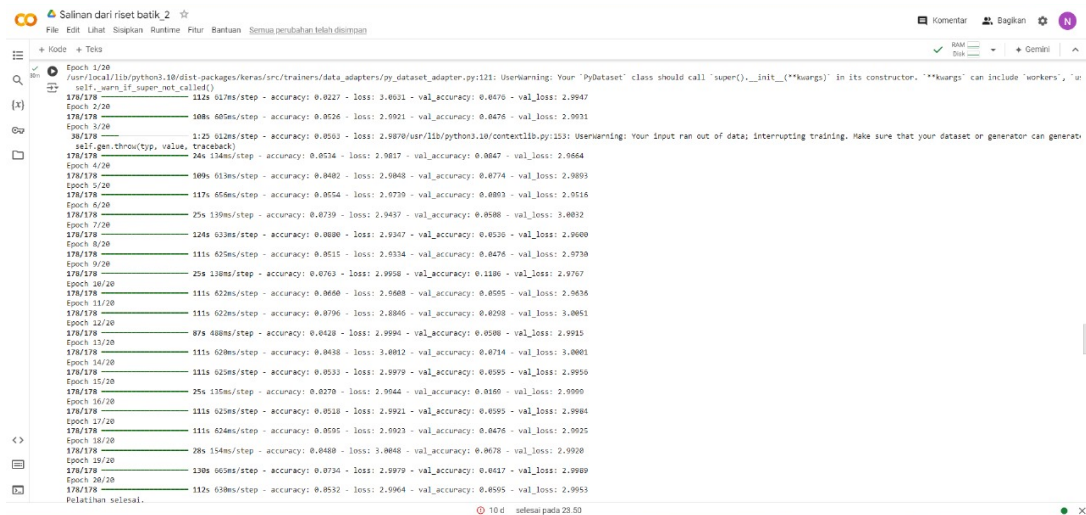
Pada gambar 4 menunjukkan bahwa hasil dari proses pelatihan model *Support Vector Machine* (SVM) untuk klasifikasi gambar batik, di mana akurasi model pada data validasi adalah 0.1975 atau sekitar 19.75%. Ini berarti model hanya berhasil mendapatkan akurasi sekitar 19.75% dari total 1200 data train.

Adapun ringkasan dari arsitektur CNN yang digunakan algoritma CNN dalam penelitian ini. Total dari 2000 data batik 60% untuk data pelatihan, 20% untuk data validasi dan 20% lagi untuk data pengujian yang digunakan dalam analisis perbandingan ini. Data gambar batik motif Nusantara sebanyak 1.200 gambar motif batik yang digunakan untuk pelatihan, data gambar motif batik sebanyak 400 gambar untuk data validasi dan data gambar motif batik Nusantara sebanyak 400 digunakan untuk pengujian. Data pelatihan dibagi menjadi 20 epoch. Sementara itu, grafik pelatihan dengan data 20 epoch ditampilkan pada gambar dibawah ini :

Tabel 3. Model Sequential

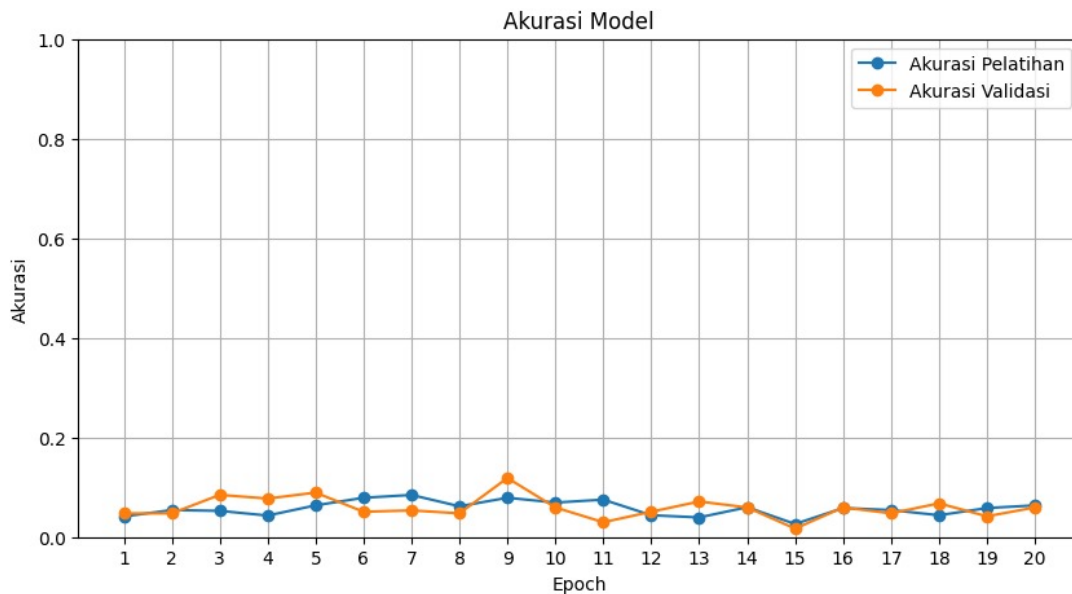
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 32)	896
max_pooling2d (MaxPooling2D)	(None, 74, 74, 32)	0
conv2d_1 (Conv2D)	(None, 72, 72, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 64)	0
conv2d_2 (Conv2D)	(None, 34, 34, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 128)	0
flatten (Flatten)	(None, 36992)	0
dense (Dense)	(None, 512)	18,940,416
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 20)	10,260

Bedasarkan Tabel 3 klasifikasi data batik ini terdiri dari beberapa lapisan, termasuk Conv2D, MaxPooling2D, Flatten, Dense, dan Dropout, dengan total 19.843.924 parameter yang seluruhnya dapat dilatih, dirancang untuk mengenali pola-pola visual pada gambar batik dan mengklasifikasikannya ke dalam 20 kategori.



Gambar 5. Hasil Pelatihan / Epoch

Bedasarkan gambar 5 model pelatihan untuk mengklasifikasikan batik dilakukan secara berkala dengan menghitung langkah per epoch (steps per epoch), dengan batch size sebesar 3, total data pelatihan sebanyak 536, dan total data validasi sebesar 170, kita dapat menentukan jumlah batch yang digunakan pada setiap epoch selama pelatihan. Berdasarkan gambar 5, terdapat tiga epoch dengan hasil tertinggi, yaitu pada Epoch 11 dengan akurasi 0,0879 dan loss 2,884, diikuti oleh Epoch 6 dengan akurasi 0,0873 dan loss 2,943, serta Epoch 20 dengan akurasi 0,0853 dan loss 2,996. Sementara itu, epoch dengan hasil terendah adalah Epoch 1 dengan akurasi 0,0227 dan loss 3,063. Dari data tersebut, dapat disimpulkan bahwa tiga epoch tertinggi menunjukkan akurasi terbaik dan loss terendah pada data validasi, yang mencerminkan peningkatan performa model seiring berjalannya pelatihan.



Gambar 6. Hasil Pelatihan / Epoch dalam bentuk grafik

Pada Gambar 6 menunjukkan perbandingan akurasi pelatihan (biru) dan akurasi validasi (oranye) dari model selama 20 epoch. Kedua akurasi tampak sangat rendah dan mendatar, dengan nilai di bawah 0.2 sepanjang proses pelatihan, yang menandakan bahwa model belum mencapai performa yang baik, baik pada data pelatihan maupun data validasi. Belum ada peningkatan yang signifikan pada akurasi di kedua set data, yang menunjukkan bahwa model belum berhasil belajar dengan baik dari data yang diberikan.

4. DISKUSI

Berdasarkan hasil pengujian yang telah dilakukan, maka dapat disimpulkan bahwa analisis Perbandingan motif batik Nusantara menggunakan metode *Convolutional Neural Network* dengan hasil akurasi dibawah 0,2 yaitu sekitar 0.08 atau hanya memiliki nilai akurasi sebesar 8% dan *Support Vector Machine* (SVM) dengan hasil akurasi data validasi adalah 0.1975 atau sekitar 19.75% yang diperoleh dari pengujian terhadap 2000 dataset batik, 60% untuk data pelatihan, 20% untuk data validasi dan 20% sebagai data pengujian. Berdasarkan hasil penelitian yang telah dilakukan metode SVM mengklasifikasikan motif batik lebih baik dibandingkan model CNN dalam dataset yang digunakan, hal ini sesuai dengan penelitian terdahulu dengan judul “ ANALISIS PERBANDINGAN ALGORITMA CNN DAN SVM PADA KLASIFIKASI EKSPRESI WAJAH “ yang diterbitkan oleh I Putu Agus Aryawan, I Nyoman Purnama dan Ketut Queena Fredlina yang menyatakan bahwa metode SVM lebih baik dibandingkan metode CNN [15], namun penelitian tersebut berbanding terbalik dengan hasil penelitian dengan judul “ Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah ” yang dibuat oleh Bobby Kurniadi W, Hariyanto Prasetyo, Ghifari Ahmad L, Bagas Aditya Wibisono, dan Desta Sandya Prasvita yang menyatakan bahwa metode metode CNN memiliki Tingkat akurasi yang lebih baik dibandingkan metode SVM [16].

5. KESIMPULAN

Penelitian terhadap 2000 data batik dengan, 60% untuk data pelatihan, 20% untuk data validasi dan 20% untuk data pengujian terhadap pelatihan model dengan dilakukannya 20 epoch/pengecekan pada metode CNN. Berdasarkan hasil penelitian yang telah dilakukan dapat diketahui bahwa metode SVM memiliki Tingkat akurasi yang lebih tinggi dibandingkan metode CNN, namun untuk penelitian kali ini kedua metode tersebut mendapatkan tingkat akurasi yang rendah dikarenakan tidak memiliki dataset yang banyak.

Penelitian ini masih memiliki batasan terhadap jumlah data dan variasi data batik. Penelitian selanjutnya dapat dilakukan dengan menggunakan lebih banyak data dan variasi data yang lebih tinggi untuk meningkatkan akurasi sistem klasifikasi batik sehingga masyarakat dapat lebih mengenal keragaman batik dari Indonesia dan penelitian selanjutnya dapata mampu membuat penambahan melalui segi kamera agar dalam melakukan pendeteksian lebih fleksibel.

DAFTAR PUSTAKA

- [1] M. Larasati, "Pelestarian Budaya Batik Nusantara Sebagai Identitas Kultural Melalui Pameran di Mosium Batik Pekalongan pada Masa Covid-19," 2021.
- [2] R. D. F. Ivona Savitri, "Penerapan Motif Busana Etnik Aceh pada Butik Vinnel Galery," *Jurnal Ilmiah Mahasiswa Program Studi Pendidikan Kesejahteraan Keluarga FKIP USK*, 2019.
- [3] D. A. W. Natalia, "Batik Bali Pada Industri Sari Amerta Batik Collection di Desa BatuBulan Kecamatan Sukawati Kabupaten Gianyar.," 2018.
- [4] J. Wu, "Introduction to Convolutional Neural Networks," *Introduction to Convolutional Neural Networks*, pp. 1–31., 2017.
- [5] P. S. d. R. A. P. 3. Dwi Puji Prabowo1, "PENGENALAN CITRA BATIK MENGGUNAKAN FITUR FRAKTAL BERDASARKAN METODE SUPPORT VECTOR," *JURNAL INFORMATIKA UPGRIS*, pp. 61-65, 2022.
- [6] F. Parapat, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *J.PTIK. Vol.2, No.10*, 2018.
- [7] E. N. A. a. H. Harintaka, "Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi," 2018.
- [8] K. N. R. a. S. M. Suyanto, "Deep Learning Modernisasi Machine Learning untuk Big Data," in *DEEP LEARNING Modernisasi Machine Learning untuk Big Data, Informatika,*, pp. 33-41, 2019.
- [9] X. X. N. a. C. Y. Suen, "'A Novel Hybrid CNN-SVM Classifier For Recognizing,'" 2011.
- [10] R. R. Saragih, "Pemrograman dan Bahasa Pemrograman.," *STMIK-STIE Mikroskil,* 2016.
- [11] R. S. Das, "TensorFlow: Revolutionizing Large-Scale Machine Learning in Complex Semiconductor Design.," *International Journal of Computing and Engineering*, 5(3), pp. 1-9, 2024.
- [12] F. V. G. G. A. M. V. T. B. G. O. .. & D. É. Pedregosa, "Scikit-learn: Machine learning in Python. the Journal of machine Learning research.," pp. 2825-2830., 2011.
- [13] F. Chollet, "'Keras: Deep learning library for theano and tensorflow.,"" 2015.
- [14] G. Bradski, "The OpenCV Library. Dr. Dobb's Journal of Software Tools.," 2000.
- [15] P. A. A. 1. N. P. 2. Q. Fredlina3), "Analisis Perbandingan Algoritma CNN dan SVM pada Klasifikasi Ekspresi Wajah," pp. 399-407, 2023.
- [16] B. K. W.1, "Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah," pp. 1-11, 2021.

SENASTIKA Universitas Malikussaleh

DIAGNOSIS PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

David Fadlianda¹, Aditya Prananto², Cut Anggel Eriska³, Syifa Anjanira⁴, Nada Syadzwin⁵, Munirul Ula⁶

^{1,2,3,4,5,6} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹david.210170272@mhs.unimal.ac.id, ²aditya.210170282@mhs.unimal.ac.id,
³cut.210170277@mhs.unimal.ac.id, ⁴syifa.210170270@mhs.unimal.ac.id,
⁵nada.210170276@mhs.unimal.ac.id, ⁶munirulula@unimal.ac.id

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian global, dengan angka kejadian yang terus meningkat akibat perubahan gaya hidup. Permasalahan utama dalam diagnosa penyakit jantung terletak pada kompleksitas dan keragaman data medis yang tersedia dan sering kali terdiri dari berbagai variabel seperti tekanan darah, kadar kolesterol, riwayat kesehatan keluarga, dan faktor-faktor lainnya. Oleh karena itu, metode pembelajaran mesin seperti *SVM* sangat diperlukan untuk membantu mengolah data dengan efisien dan meningkatkan akurasi prediksi. Metode yang digunakan dalam penelitian ini meliputi beberapa tahapan, yaitu pengumpulan data medis, pra-pemrosesan data untuk mengatasi ketidakseimbangan dan kebisingan data, penerapan algoritma *SVM* untuk membangun model prediksi, serta evaluasi hasil dengan menggunakan dataset yang telah disediakan. Dataset yang digunakan mencakup berbagai parameter medis yang relevan. Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang akurat menggunakan metode pembelajaran mesin, khususnya algoritma Support Vector Machine (SVM), dalam mendeteksi risiko penyakit jantung. Penelitian ini bertujuan untuk mengatasi permasalahan kompleksitas dan keragaman data medis. Evaluasi dilakukan dengan menggunakan metrik-metrik seperti akurasi, precision, recall, dan F1-Score untuk menilai kinerja model dalam memprediksi risiko penyakit jantung. Hasil penelitian menunjukkan bahwa model *SVM* mampu mencapai akurasi sebesar 86,89%, dengan nilai precision, recall, dan F1-Score masing-masing sebesar 0,88. Ini menunjukkan bahwa model *SVM* memiliki kemampuan yang baik dalam mengklasifikasikan pasien dengan risiko tinggi maupun rendah terhadap penyakit jantung. Selain itu, model ini menunjukkan kinerja yang stabil dan dapat diandalkan, baik untuk pasien yang sudah terdiagnosis maupun yang belum menunjukkan gejala yang signifikan. Dengan demikian, penggunaan algoritma *SVM* dalam diagnosis penyakit jantung terbukti efektif, menawarkan solusi yang signifikan dalam membantu dokter dan tenaga medis untuk melakukan deteksi dini serta intervensi yang lebih tepat waktu.

Kata kunci: *Algoritma, Medis, Machine Learning, Penyakit Jantung, Support Vector Machine.*

Abstract

Cardiovascular disease is one of the leading causes of death globally, with the incidence rate continuing to increase due to lifestyle changes. The main problem in diagnosing heart disease lies in the complexity and diversity of available medical data, which often consists of various variables such as blood pressure, cholesterol levels, family health history, and other factors. Therefore, machine learning methods such as *SVM* are needed to help process data efficiently and improve prediction accuracy. The method used in this study includes several stages, namely medical data collection, data pre-processing to overcome data imbalance and noise, application of the *SVM* algorithm to build a prediction model, and evaluation of the results using the provided dataset. The dataset used includes various relevant medical parameters. The purpose of this study is to develop an accurate prediction model using machine learning methods, especially the Support Vector Machine (*SVM*) algorithm, in detecting the risk of heart disease. This study aims to overcome the problems of complexity and diversity of medical data. The evaluation was carried out using metrics such as accuracy, precision, recall, and F1-Score to assess the performance of the model in predicting the risk of heart disease. The results showed that the *SVM* model was able to achieve an accuracy of 86.89%, with precision, recall, and F1-Score values of 0.88 each. This shows that the *SVM* model has good ability in classifying patients with high and low risk of heart disease. In addition, this model shows stable and reliable performance, both for patients who have been diagnosed and those who have not shown significant symptoms. Thus, the use of the *SVM* algorithm in the diagnosis of heart disease has proven effective, offering a significant solution in helping doctors and medical personnel to carry out early detection and more timely intervention.

Keywords: *Algorithm, Medical, Machine Learning, Cardiovascular disease, Support Vector Machine.*

1. PENDAHULUAN

Penyakit jantung adalah penyebab utama kematian di seluruh dunia, dan kejadiannya terus meningkat seiring dengan perubahan gaya hidup dan kebiasaan makan masyarakat. Jantung adalah salah satu organ terpenting manusia, dan perannya adalah memompa darah ke seluruh bagian tubuh. Hal tersebut membuat jantung memiliki peranan yang sangat penting untuk tubuh manusia, oleh sebab itu penderita penyakit jantung akan merasakan ketidaknyamanan pada tubuh mereka. Kemudian penyakit jantung juga dapat menyebabkan penghambatan dari fungsi jantungnya seperti kardiovaskular, jantung koroner, bahkan yang paling parahnya dapat menyebabkan gagal jantung [1]. Penyakit jantung adalah penyakit mematikan yang menyerang hampir 12 juta orang di seluruh dunia. Oleh karena itu pemeriksaan sedari dini merupakan upaya yang sangat penting untuk dilakukan, dan diagnosa penyakit jantung menjadi salah satu hal yang sangat beresiko dikarenakan memiliki ikatan yang saling berkegantungan dari berbagai faktor [2]. Meskipun kemajuan dalam teknologi medis memungkinkan pengumpulan data yang lebih banyak dan beragam, masih terdapat kesulitan dalam mengolah dan menganalisis data tersebut untuk mengidentifikasi individu yang berisiko tinggi. Masalah utama yang dihadapi adalah bagaimana memanfaatkan data medis yang tersedia seperti riwayat kesehatan, faktor risiko, dan hasil pemeriksaan untuk memberikan gambaran yang jelas tentang risiko penyakit jantung secara keseluruhan.

Dalam konteks ini, algoritma *machine learning*, dengan metode *Support Vector Machine (SVM)*, menawarkan solusi yang menjanjikan. *SVM* dapat menganalisis data berukuran besar dan kompleks, membantu dalam mengidentifikasi pola yang mungkin tidak terlihat melalui metode tradisional. Dengan kemampuannya untuk menemukan batas optimal antara kelas-kelas data, *SVM* dapat digunakan untuk memprediksi risiko penyakit jantung secara lebih akurat.

Metode ini menggunakan prinsip dasar pengklasifikasi linier, yaitu kasus pengklasifikasi yang dapat diklasifikasikan secara linier. *SVM* dikembangkan untuk menangani masalah nonlinier dengan menerapkan teori kernel pada ruang kerja yang besar. Hyperplanes muncul pada ruang dengan dimensi besar, yang dapat memaksimalkan selisih (margin) antar dataset [4].

Penelitian sebelumnya yang dikembangkan oleh Herliani Hasanah dan Nurmaalityari, “Perbandingan tingkat akurasi support vector machine (SVM) dan algoritma C45 dalam memprediksi penyakit jantung”, menunjukkan bahwa algoritma SVM merupakan algoritma yang akurat dan akurat digunakan untuk memprediksi penyakit jantung hasilnya menunjukkan hal itu skor akurasi 87%. Sedangkan untuk algoritma C45 memiliki nilai akurasi sebesar 82% dalam menyelesaikan kasus klasifikasi penyakit jantung. Secara keseluruhan Algoritma SVM mempunyai nilai keakuratan yang lebih tinggi jika dibandingkan dengan algoritma C45 [5]. Penelitian terdahulu lainnya dengan menggunakan subjek penelitian yang berbeda dilakukan pada jurnal penelitian berjudul “Machine Learning Breast Cancer Diagnosis Menggunakan Algoritma Support Vector Machine” oleh Chalifa Chazar dan Bagus Erawan Widhiaputra dan SVM yang digunakan dalam penelitian ini Pembentukan sistem aplikasi ML menggunakan algoritma adalah dibahas. Diagnosis kanker payudara. Dataset yang digunakan adalah dataset Wisconsin Breast Cancer (Diagnosis). Dan hasil penelitian ini menunjukkan bahwa aplikasi ML ini dapat mendiagnosis kanker payudara dan mengambil keputusan prediktif berdasarkan dua kemungkinan: sel hidup dalam keadaan ganas atau jinak. Aplikasi ini juga dapat digunakan oleh petugas laboratorium untuk membuat diagnosis berdasarkan hasil FNA dari studi biopsi [6].

Dengan itu, tujuan pada penelitian ini untuk mengimplementasikan algoritma SVM guna memprediksi risiko penyakit jantung berdasarkan dataset yang telah disediakan. Penelitian ini bertujuan tidak hanya untuk menghasilkan model prediktif yang akurat, tetapi juga untuk memberikan wawasan mengenai faktor-faktor yang signifikan dalam pengembangan penyakit jantung. Penelitian ini dilaksanakan menggunakan *Google Colabs*, yang memungkinkan peneliti untuk mengakses lingkungan pemrograman yang interaktif dan efisien dalam mengolah data serta menerapkan algoritma. Melalui langkah-langkah seperti pengumpulan dan pra-pemrosesan data, penerapan algoritma SVM, serta evaluasi hasil, kami berharap dapat menyajikan analisis yang mendalam tentang variabel-variabel yang berpengaruh.

Secara Signifikansi penelitian ini terletak pada potensi kontribusinya terhadap bidang kesehatan, khususnya dalam meningkatkan akurasi diagnosis penyakit jantung dan membantu profesional medis untuk pengambilan keputusan yang lebih efektif dan simple. Dengan menyediakan model prediktif berbasis data, penelitian ini diharapkan dapat mendukung upaya pencegahan penyakit jantung dan memajukan kualitas layanan kesehatan di masyarakat. Implementasi algoritma SVM pada penelitian ini juga akan menunjukkan bagaimana teknologi machine learning dapat diintegrasikan dalam praktik medis untuk mencapai hasil yang lebih baik.

2. TINJAUAN PUSTAKA

Tinjauan ini berperan penting dalam memposisikan penelitian secara tepat, memastikan penggunaan metodologi yang valid, dan memperkuat argumen penggunaan machine learning metode SVM untuk diagnosis penyakit jantung. Analisis literatur akan mengidentifikasi celah pengetahuan, mendorong inovasi, dan berkontribusi pada peningkatan akurasi dalam deteksi dan klasifikasi penyakit jantung menggunakan dataset pasien.

2.1 Machine Learning

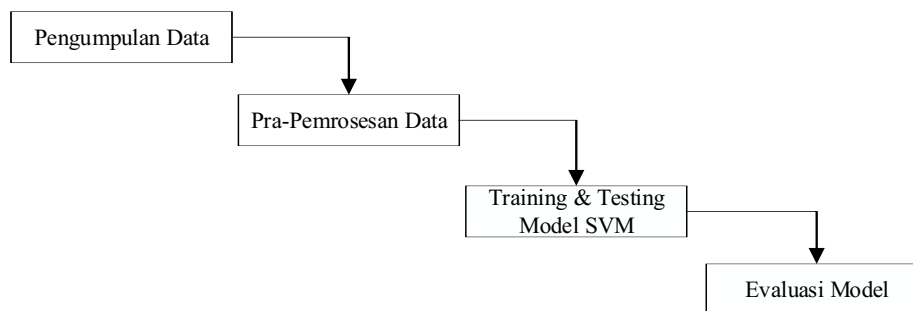
Machine learning dapat disebut dengan sebuah cabang kecerdasan buatan yang fokus terhadap pengembangan sistem yang dapat memahami dan berkembang secara mandiri. Inti dari teknologi ini adalah kemampuannya untuk mengidentifikasi pola kompleks dalam kumpulan data besar, tanpa perlu diprogram secara eksplisit untuk setiap skenario [7]. Dengan memanfaatkan algoritma canggih dan model statistik, sistem machine learning ini dapat membuat keputusan atau prediksi berdasarkan pengalaman sebelumnya, mirip dengan proses pembelajaran manusia. Seiring berjalannya waktu dan semakin banyaknya data yang diproses, kinerja sistem ini terus meningkat, memungkinkan aplikasinya dalam berbagai bidang termasuk dalam diagnosis dataset pasien penyakit jantung [8].

2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan sebuah algoritma supervised machine learning yang digunakan dalam sebuah tugas klasifikasi, yang mengklasifikasikan data menjadi dua kelas: +1 dan -1. SVM bekerja dengan mencari hyperplane optimal yang memaksimalkan jarak antara dua kelas dan memberikan pemisahan paling akurat. Algoritma ini dapat mengatasi nonlinier dengan menggunakan trik kernel sehingga sangat efektif terutama dalam menangani dataset yang kompleks dan besar dengan banyak fitur, menjadikannya pilihan populer dalam berbagai aplikasi klasifikasi [9].

3. METODE PENELITIAN

Melakukan penelitian yang berbasis data dan informasi objektif sangat penting sebagai landasan dalam penelitian tersebut. Dengan adanya data yang akurat, diharapkan hasil penelitian akan berkualitas tinggi [10]. Berikut adalah proses penelitian yang dapat digunakan dalam merangkum informasi tentang penelitian ini:



Gambar 1. Metode Penelitian

3.1 Pengumpulan Data

Dalam pengembangan sistem prediksi penyakit jantung, penelitian ini dilaksanakan melalui beberapa langkah. Pertama, data diperoleh dari sumber yang kredibel yang berupa file *Comma Separated Values (CSV)*. Dataset ini didapatkan melalui sebuah *website kaggle* [11] <https://www.kaggle.com/code/rafiromolo/prediksi-penyakit-jantung>. Dari hasil dataset ini didapatkan total hasil sebanyak 303 pasien. Setelah itu, data tersebut diproses untuk menjamin kualitas dan konsistensinya. Implementasi sistem ini dilakukan menggunakan bahasa pemrograman *Python* dengan sebuah platform *Google Collab* yang berbasis cloud untuk menulis tiap kode yang dibutuhkan pada algoritma *Support Vector Machine (SVM)* yang digunakan.

3.2 Pra-Pemrosesan Data

Pra-processing merupakan tahap yang sangat penting sebelum data bisa digunakan [12] untuk melatih algoritma *Support Vector Machine (SVM)*. Dengan metode ini dimana terdapat proses dalam memisahkan fitur dan target, membagi dataset menjadi training dan testing dan scaling data. Tujuannya adalah memastikan bahwa data dalam format yang benar, bersih, dan siap untuk digunakan oleh model.

3.3 Training & Testing Model

Proses Training Model, mempelajari data yang telah diproses menggunakan data latihan. Sementara itu, Testing Model adalah proses di mana model membuat prediksi berdasarkan data uji. Kedua proses ini sangat penting karena melalui pemrosesan dataset secara otomatis, *SVM* dapat menghasilkan model yang paling optimal [13].

3.4 Evaluasi Model

Pengujian model diperlukan untuk mengevaluasi kinerja metode support vector machine (SVM). Dalam studi ini, evaluasi didasarkan pada akurasi, yaitu metrik yang mengukur seberapa akurat model memprediksi hasil yang benar berdasarkan data.[14]. Akurasi ini akan menunjukkan tingkat keberhasilan model dalam memprediksi label

yang benar. Dalam pengujian metode SVM ini, yang juga akan memberikan persentase terhadap performa *confusion matrix* dari setiap rumus yang sudah ditentukan melalui metode SVM ini. Persamaan rumus tersebut dapat dilihat seperti berikut:

1. Akurasi merupakan suatu nilai rasio information tweet yang nyata saat dideteksi di dalam information pengujian. Dengan kata lainnya akurasi ini merupakan nilai yang secara langsung menunjukkan tingkat kedekatan antara suatu nilai prediksi sistem dengan suatu nilai prediksi manusia [15]. Nilai akurasi dapat dicari dengan persamaan sebagai berikut :

$$Accuracy = \left(\frac{TP + TN}{TP + FP + TN + FN} \right) \times 100 \% \quad (1)$$

2. Precision mengukur tingkat akurasi antara data yang diinginkan dengan hasil prediksi yang dihasilkan oleh model. Precision adalah rasio dari prediksi positif yang benar dibandingkan dengan seluruh prediksi yang dikategorikan sebagai positif. Konsep ini menjelaskan seberapa tepat hasil prediksi model terhadap data yang diinginkan, menunjukkan sejauh mana prediksi positif yang dibuat oleh model sesuai dengan kenyataan [16]. Nilai precision dapat dicari menggunakan persamaan dibawah ini :

$$Precision = \left(\frac{TP}{TP + FP} \right) \times 100 \% \quad (2)$$

3. Recall membantu kemampuan model untuk mengidentifikasi ulang informasi yang relevan. Recall atau sensitivitas dipergunakan untuk mendapat hasil perbandingan jumlah suatu prediksi positif yang akurat dibandingkan dengan jumlah semua kelas yang bernilai positif [17]. Adapun nilai recall dapat dicari menggunakan persamaan dibawah ini :

$$Recall = \left(\frac{TP}{TP + FN} \right) \times 100 \% \quad (3)$$

4. F1-Score merepresentasikan rata-rata harmonik dari precision dan recall yang telah diberi bobot. Dengan menggabungkan nilai precision dan recall, F1-Score menghasilkan satu metrik tunggal yang mencerminkan keseimbangan antara keduanya, memberikan gambaran yang lebih komprehensif mengenai kinerja model dalam prediksi positif. [18]. Nilai F1-Score dapat dicari menggunakan persamaan berikut :

$$F1 - Score = \left(\frac{2 \times precision \times recall}{precision + recall} \right) \times 100 \% \quad (4)$$

Untuk mengukur akurasi, presisi, recall dan F1-Score dapat menggunakan *confusion matrix*. *Confusion matrix* merupakan alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah dari nilai presisi klasifikasi terhadap kelas menggunakan algoritma yang digunakan [19]. Berikut telah dipresentasikan bentuk *confusion matrix* dari tabel dibawah ini:

Tabel 1. Bentuk Confusion matrix dari dua kelas

Confusion Matrix		Nilai Sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

4. ANALISIS DAN PEMBAHASAN

4.1 Analisis Atribut

Dari dataset ini yang akan digunakan untuk penelitian ini, merupakan sumber dari Kaggle, terdapat total 304 pasien dengan 14 atribut. Dari 14 atribut tersebut, 13 atribut digunakan untuk menggambarkan jenis penyakit yang diderita oleh setiap pasien, sementara 1 atribut berfungsi sebagai “target” untuk menunjukkan status kesehatan pasien terkait penyakit jantung. Atribut target ini memiliki 2 nilai: nilai 1 menandakan bahwa pasien memiliki penyakit jantung, sedangkan nilai 0 menandakan bahwa pasien tidak memiliki penyakit jantung. Data atribut yang lebih rinci dapat dijelaskan pada tabel berikut:

Tabel 2. Atribut Prediksi

No	ATRIBUT	Tipe Data	Keterangan
1	Age	int64	Usia
2	Sex	int64	Jenis Kelamin

3	Cp	int64	Sakit Dada
4	Trestbps	int64	Tekanan Darah
5	Chol	int64	Kolestrol
6	Fbs	int64	Gula Darah
7	Restecg	int64	Kondisi Jantung
8	Thalach	int64	Denyut Jantung
9	Exang	int64	Exercise Induced Angina
10	Oldpeak	float64	Uji Stress Jantung
11	Slope	int64	Uji Kemiringan Jantung
12	Ca	int64	Pembuluh darah koroner
13	Thal	int64	Deteksi Fungsi Jantung
14	Target	int64	Deteksi Penyakit Jantung

4.2 Preview Dataset

Untuk memberikan gambaran awal mengenai dataset yang digunakan dalam penelitian ini, penting untuk menampilkan setidaknya 5 data teratas agar proses *pre-processing* data dapat dilakukan dengan lebih mudah sebelum melakukan pemodelan dan visualisasi pemisahan. Tabel berikut menyajikan data teratas dari dataset, yang memberikan insight awal mengenai karakteristik data yang akan digunakan dalam analisis lebih lanjut. Dengan meninjau data ini, peneliti dapat memahami struktur dataset dan mengidentifikasi langkah-langkah yang diperlukan dalam *pre-processing* sebelum melanjutkan ke tahap pemodelan.

Tabel 3. Hasil Processing

Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal	Target
63.0	1.0	3.0	145.0	233.0	1.0	0.0	150.0	0.0	2.3	0.0	0.0	1.0	1.0
37.0	1.0	2.0	130.0	250.0	0.0	1.0	187.0	0.0	3.5	0.0	0.0	2.0	1.0
41.0	0.0	1.0	130.0	204.0	0.0	0.0	172.0	0.0	1.4	2.0	0.0	2.0	1.0
56.0	1.0	1.0	120.0	236.0	0.0	1.0	178.0	0.0	0.8	2.0	0.0	2.0	1.0
57.0	0.0	0.0	120.0	354.0	0.0	1.0	163.0	1.0	0.6	2.0	0.0	2.0	1.0

4.3 Preview Model SVM

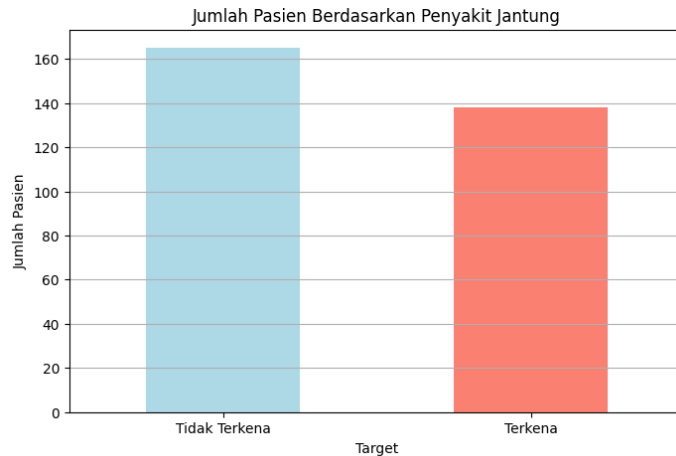
Langkah berikutnya adalah memisahkan dataset menjadi fitur (X) dan target (y), sehingga model dapat fokus pada variabel input untuk memprediksi hasil. Dengan visualisasi data menggunakan matplotlib, pola-pola penting yang mempengaruhi performa model dapat diidentifikasi. Setelah itu, model SVM dilatih menggunakan data tersebut, di mana pemahaman yang baik terhadap distribusi data sangat penting untuk memilih kernel yang tepat, dalam hal ini kernel linier. Hasil keseluruhan dari klasifikasi memperlihatkan perbedaan yang jelas antara pasien yang terkena dan tidak terkena penyakit jantung, yang akan divisualisasikan dalam bentuk tabel dan gambar berikut untuk memudahkan analisis.

Tabel 4. Fitur (X)

Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal
63.0	1.0	3.0	145.0	233.0	1.0	0.0	150.0	0.0	2.3	0.0	0.0	1.0
37.0	1.0	2.0	130.0	250.0	0.0	1.0	187.0	0.0	3.5	0.0	0.0	2.0
41.0	0.0	1.0	130.0	204.0	0.0	0.0	172.0	0.0	1.4	2.0	0.0	2.0
56.0	1.0	1.0	120.0	236.0	0.0	1.0	178.0	0.0	0.8	2.0	0.0	2.0
57.0	0.0	0.0	120.0	354.0	0.0	1.0	163.0	1.0	0.6	2.0	0.0	2.0

Tabel 5. Target (Y)

Target (y)
1
1
1
1
1



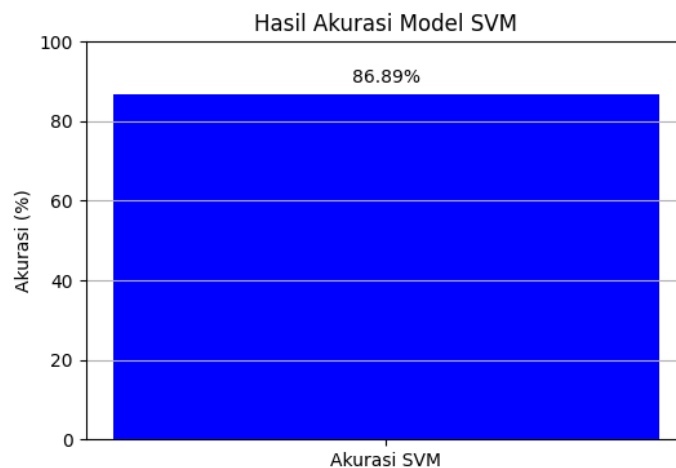
Gambar 2. Barplot Jumlah Data Pasien

Grafik di atas menampilkan jumlah pasien berdasarkan status penyakit jantung, yang dibagi menjadi dua kategori: "Tidak Terkena" dan "Terkena." Dari grafik ini, dapat diketahui bahwa jumlah pasien yang tidak terkena penyakit jantung (sekitar 160) dan lebih besar dibandingkan dengan pasien yang terkena penyakit jantung (sekitar 140).

Dalam konteks model SVM untuk diagnosis penyakit jantung, ketidakseimbangan ini dapat membuat model lebih akurat dalam memprediksi pasien yang tidak terkena, namun kurang sensitif terhadap pasien yang sebenarnya terkena penyakit. Untuk mengatasi ini, teknik pada SVM dapat diterapkan dalam evaluasi seperti precision, recall, dan F1-score yang penting digunakan untuk memastikan model tidak bias dan mampu mendeteksi risiko dengan baik, terutama pada pasien yang terkena penyakit jantung.

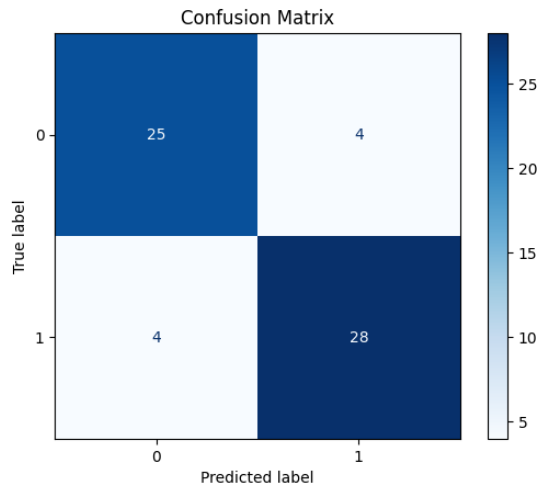
4.4 Hasil Pemodelan Data

Model SVM dilatih menggunakan kernel linier, dan akurasi model diukur serta divisualisasikan dalam bentuk diagram batang dengan nilai akurasi mencapai 86,89%. Visualisasi ini membantu memberikan gambaran yang lebih jelas mengenai performa model dalam mengklasifikasikan data secara akurat. Hasil akurasi model ditampilkan melalui gambar grafik seperti berikut, yang memperlihatkan perbandingan akurasi dengan metrik lainnya untuk evaluasi yang lebih komprehensif.



Gambar 3. Tingkat Akurasi Data

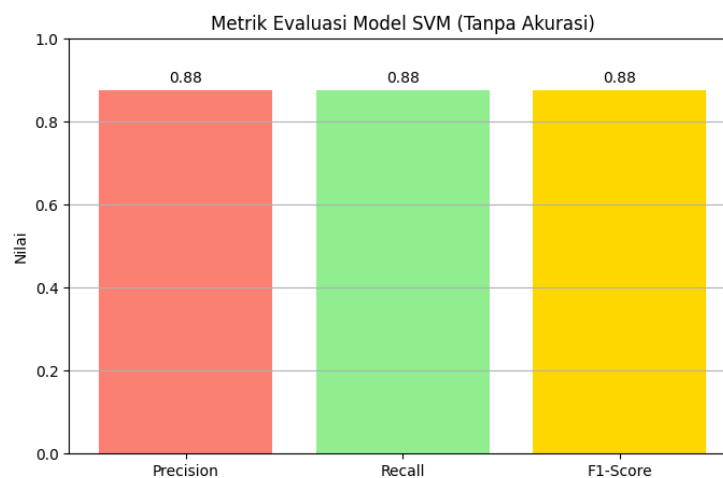
Sementara itu, dengan memanfaatkan confusion matrix, kita dapat menilai kinerja model klasifikasi, khususnya pada masalah klasifikasi biner atau multi-kelas. Confusion matrix menampilkan hasil prediksi model dengan membandingkan nilai aktual dengan nilai prediksi yang dihasilkan. Tabel ini mengategorikan hasil prediksi ke dalam empat kategori: True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN). Dari informasi ini, kita dapat menghitung berbagai metrik evaluasi seperti precision, recall, dan F1-score, yang memberikan pemahaman yang jelas mengenai seberapa baik model dalam memprediksi kelas dengan akurat serta meminimalkan kesalahan.

Gambar 4. *Confusion Matrix*

Berikut adalah penjelasan dari setiap elemen pada matrix:

1. True Negatives (TN): Sebanyak 25 sampel kelas 0 diprediksi dengan benar sebagai kelas 0.
2. False Positives (FP): Sebanyak 4 sampel kelas 0 diprediksi secara salah sebagai kelas 1.
3. False Negatives (FN): Sebanyak 4 sampel kelas 1 diprediksi secara salah sebagai kelas 0.
4. True Positives (TP): Sebanyak 28 sampel kelas 1 diprediksi dengan benar sebagai kelas 1.

Dengan demikian, melalui confusion matrix, hasil prediksi model dapat dievaluasi untuk mendapatkan nilai metrik penting seperti precision, recall, dan F1-score. Berdasarkan evaluasi tersebut, model menghasilkan nilai precision sebesar 0,88, recall sebesar 0,88, dan F1-score sebesar 0,88. Nilai-nilai ini menunjukkan bahwa model mampu melakukan prediksi dengan konsistensi yang baik, seperti yang terlihat pada gambar matrix berikut:

Gambar 5. *Metrix Evaluasi Model*

Dalam situasi tertentu, model prediksi dapat menunjukkan keseimbangan yang baik antara hasil positif dan negatif, terutama ketika jumlah True Positives (TP), False Positives (FP), dan False Negatives (FN) hampir sama atau sangat kecil. Dalam kondisi ini, Precision, Recall, dan F1-Score cenderung setara, menunjukkan bahwa model dapat secara efektif mengidentifikasi kasus positif (precision) sekaligus mendeteksi sebagian besar kasus positif yang ada (recall). Hal ini sangat penting dalam aplikasi seperti diagnosa penyakit jantung, di mana keseimbangan antara prediksi yang benar dan kesalahan prediksi harus dijaga untuk meminimalkan risiko.

5. DISKUSI

5.1 Ketidakseimbangan

Distribusi data pada penelitian ini menunjukkan adanya ketidakseimbangan antara jumlah pasien yang tidak terkena penyakit jantung (160 pasien) dan yang terkena penyakit jantung (144 pasien) dari total 304 pasien. Ketidakseimbangan data ini memberikan tantangan tersendiri dalam membangun model prediksi yang akurat menggunakan algoritma Support Vector Machine (SVM). Secara umum, ketidakseimbangan seperti ini dapat

menyebabkan model lebih cenderung memprediksi kategori mayoritas, yaitu pasien yang tidak terkena penyakit jantung, dan mungkin kurang mampu mendeteksi pasien yang benar-benar berisiko. Hal ini dapat terlihat pada metrik seperti precision dan recall, di mana recall untuk pasien yang terkena penyakit jantung mungkin lebih rendah karena ketidakseimbangan ini.

5.2 Kinerja Model SVM

Model SVM yang diterapkan pada dataset ini menunjukkan hasil akurasi yang cukup baik, dengan nilai akurasi sebesar 86,89%. Namun, akurasi saja tidak cukup untuk mengevaluasi kinerja model, terutama dengan adanya ketidakseimbangan data. Precision untuk pasien yang terkena penyakit jantung mencapai 0,88, yang menunjukkan bahwa model cukup baik dalam menghindari prediksi false positive (memastikan bahwa pasien yang diprediksi terkena benar-benar berisiko). Namun, recall juga harus diperhatikan, karena model bisa saja gagal mendeteksi sejumlah pasien yang terkena penyakit jantung (false negative). Nilai recall yang lebih rendah akan menunjukkan bahwa model masih bisa melewatkan pasien yang benar-benar berisiko. Dalam hal ini, metrik F1-Score (kombinasi antara precision dan recall) digunakan untuk memberikan gambaran keseimbangan antara kedua aspek ini, yang dapat membantu menentukan apakah model SVM sudah optimal.

5.3 Perbandingan dengan Metode Lain

Penelitian ini menggunakan algoritma SVM dengan akurasi 86,89%, yang sejalan dengan penelitian oleh Herliani Hasanah dan Nurmaalitayari yang menunjukkan akurasi SVM sebesar 87%, lebih tinggi dibandingkan algoritma C45 yang hanya mencapai 82%. Hal ini menunjukkan bahwa SVM lebih unggul dalam memprediksi penyakit jantung dan pada penelitian oleh Chalifa Chazar dan Bagus Erawan Widhiaputra dalam diagnosis kanker payudara menggunakan SVM juga menunjukkan akurasi tinggi dalam membedakan sel ganas dan jinak. Kedua penelitian ini menegaskan bahwa SVM lebih efektif dibandingkan metode lain seperti C45 dalam menangani data medis yang kompleks, baik untuk penyakit jantung maupun kanker.

6. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan metode Support Vector Machine (SVM) dalam diagnosa penyakit jantung terbukti sangat efektif, dengan tingkat akurasi mencapai 86,89%. Pada penelitian ini, model SVM diuji dengan menggunakan pembagian dataset di mana 80% digunakan sebagai data training dan 20% sebagai data testing. Selain akurasi yang cukup tinggi, model ini juga menunjukkan performa yang konsisten dengan nilai precision, recall, dan F1-Score yang sama, yaitu 0,88. Hal ini menunjukkan bahwa model SVM mampu memprediksi secara akurat baik pasien yang menderita penyakit jantung maupun yang tidak, serta memberikan keseimbangan yang baik dalam mendeteksi kasus positif sekaligus meminimalkan kesalahan prediksi. Dengan demikian, hasil penelitian ini memperkuat efektivitas SVM sebagai alat bantu yang andal dalam mendiagnosa data pasien terkait penyakit jantung, memberikan kontribusi signifikan dalam bidang kesehatan untuk deteksi dini penyakit jantung.

DAFTAR PUSTAKA

- [1] E. Tasia, R. Zaid, I. Zarier, S. Kenia, and P. Loka, "Klasifikasi Penyakit Gagal Jantung Menggunakan Supervised Learning," *Sentimas*, pp. 1–7, 2023.
- [2] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [3] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics and Digital Expert (INDEX)*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [4] W. Latuny, V. O. Lawalata, D. B. Paillin, and R. Ohoirenana, "Prediksi Fitur Kemasan Produk Minyak Kayu Putih Dengan Support Vector Machine (Svm)," *ALE Proceeding*, vol. 4, pp. 76–82, 2021, doi: 10.30598/ale.4.2021.76-82.
- [5] Herliyani, Hasanah and Nurmalitasari, "Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C45 dalam Prediksi Penyakit Jantung," *Prosiding Seminar Nasional Teknologi dan Sains*, vol. 2, pp. 13–18, 2023.
- [6] C. Chazar and B. E. Widhiaputra, "INFORMASI (Jurnal Informatika dan Sistem Informasi) Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," 2020.
- [7] S. Prasetyo and T. Dewayanto, "Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan-a Systematic Literature Review," *Diponegoro Journal of Accounting*, vol. 13, no. 3, pp. 1–12, 2024, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/accounting>
- [8] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, "Implementasi Arsitektur Xception Pada Model Machine Learning Klasifikasi Sampah Anorganik," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 2, pp. 233–236, 2023, doi: 10.23960/jitet.v11i2.3034.

- [9] A. C. Nugraha and M. I. Irawan, "Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost)," *Jurnal Sains dan Seni ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.107032.
- [10] Jasmarizal, Junadhi, Rahmadden, and M. Khairul Anam, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Terhadap Produk Skincare," *Indonesian Journal of Computer Science*, vol. 13, no. 1, pp. 1438–1450, 2024, doi: 10.33022/ijcs.v13i1.3654.
- [11] J. Teknik Elektro dan Komputasi *et al.*, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 4, no. 2, pp. 207–214, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/ELKOM/article/view/7691>
- [12] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *EduTic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [13] Lukman Priyambodo *et al.*, "Klasifikasi Kematangan Tanaman Hidroponik Pakcoy Menggunakan Metode SVM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 153–160, 2022, doi: 10.29207/resti.v6i1.3828.
- [14] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [15] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.
- [16] H. Jurnal, A. Rusdy Prasetyo, and B. S. Aditya, "JURNAL ILMIAH TEKNIK MESIN, ELEKTRO DAN KOMPUTER ANALISIS PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK SISTEM DETEKSI KATARAK," vol. 3, pp. 1–10, 2023.
- [17] S. N. N. Arif, A. M. Siregar, S. Faisal, and A. R. Juwita, "Klasifikasi Penyakit Serangan Jantung Menggunakan Metode Machine Learning K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, p. 1617, Jul. 2024, doi: 10.30865/mib.v8i3.7844.
- [18] E. Hasibuan and E. A. Heriyanto, "ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER," *JTS*, vol. 1, no. 3, 2022.
- [19] C. Emilia Sukmawati Universitas Buana Perjuangan Karawang Kab Karawang, A. Rizky Pratama Universitas Buana Perjuangan Karawang Kab Karawang, and N. Nurjanah Universitas Buana Perjuangan Karawang Kab Karawang, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi ALGORITMA NAÏVE BAYES, RANDOM FOREST DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT SAPI," 2024.

SENASTIKA Universitas Malikussaleh

IMPLEMENTASI DETEKSI BERITA CLICKBAIT BERBAHASA INDONESIA MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Dila Puspita^{*1}, Rizal Tjuet Adek², Hafizh Al Kautsar Aidilof^{*3}

^{1,2,3,4} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: ¹dila.200170024@mhs.unimal.ac.id, ²rizal@unimal.ac.id, ³hafizh@unimal.ac.id

Abstrak

Berita *online* semakin populer seiring dengan peningkatan penggunaan internet di masyarakat. Namun, fenomena *clickbait* muncul ketika banyak produsen berita menyesatkan pembaca dengan judul yang tidak relevan atau sensasional. Judul-judul tersebut sering kali tidak mencerminkan isi konten secara akurat, sehingga mengecewakan pembaca dan merusak reputasi media. Penelitian ini bertujuan untuk membangun sebuah sistem yang mampu mendeteksi berita *clickbait* dan *non-clickbait* secara otomatis. Sistem ini diharapkan dapat membantu pembaca mendapatkan informasi yang lebih akurat dan relevan serta meningkatkan kualitas berita online secara keseluruhan. Metode yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM). SVM dipilih karena kemampuannya dalam mengidentifikasi *hyperplane* optimal yang memisahkan dua kelompok data, serta dapat menangani masalah *linier* dan *non-linier*. Dengan menggunakan data latih terlabel, SVM akan menemukan pola dalam fitur judul dan isi konten berita untuk melakukan klasifikasi antara *clickbait* dan *non-clickbait*. Dataset yang terdiri dari 8.660 berita telah dianotasi, dengan 6.888 berita digunakan sebagai data latih dan 1.772 berita sebagai data uji. Hasil evaluasi model menunjukkan akurasi sebesar 82%, precision 81%, recall 65%, dan F1-score 72%. Kinerja ini membuktikan bahwa metode SVM efektif dalam memisahkan berita *clickbait* dari *non-clickbait*, yang pada akhirnya dapat membantu mengurangi penyebaran informasi yang menyesatkan di platform berita *online*.

Keywords: *Berita Online, Clickbait, Non-clickbait, Klasifikasi, Support Vector Machine (SVM).*

1. PENDAHULUAN

Berita *online* semakin populer sebagai cara untuk mendapatkan informasi seiring dengan meluasnya penggunaan internet di masyarakat. Berita dapat dengan mudah diakses dan didistribusikan ke berbagai perangkat melalui internet. Namun banyak produsen berita menggunakan ini untuk menyesatkan pembaca. Fenomena menyesatkan pembaca ini terlihat antara judul berita dan konten yang disampaikan tidak relevan. Fenomena ini disebut *clickbait* [1].

Clickbait adalah jenis berita yang menggunakan judul yang sensasional atau menyesatkan untuk menarik perhatian pembaca dan meningkatkan klik. Meskipun tidak selalu mengandung informasi yang salah, *clickbait* sering kali mengecewakan pembaca karena isi kontennya yang tidak sebanding dengan ekspektasi yang dibangun oleh judulnya. Ini tidak hanya merugikan pembaca, tetapi juga dapat merusak reputasi media dan menurunkan kualitas informasi yang beredar di masyarakat. Ciri-ciri berita *clickbait* ditandai dengan penggunaan judul yang emosional, berlebihan, atau menyesatkan. Judul-judul tersebut sering kali dimaksudkan untuk menarik perhatian pembaca, terlepas dari keakuratan atau kelengkapan informasi yang disampaikan [2].

Didasarkan pada permasalahan *clickbait*, salah satu metode yang dapat digunakan dalam penelitian ini adalah menggunakan metode *Support Vector Machine*. SVM sebuah teknik yang berupaya mengidentifikasi *hyperplane* optimal yang membagi ruang input menjadi dua kelompok. SVM memiliki kemampuan untuk menangani masalah linier dan nonlinier [3]. SVM bekerja dengan mengidentifikasi *hyperplane* optimal yang membagi ruang input menjadi dua kelompok, yaitu *clickbait* dan *non-clickbait*. Dalam proses deteksi *clickbait*, SVM akan memisahkan judul dan isi konten berita yang bersifat *clickbait* dan *non-clickbait* didasarkan fitur-fitur yang disediakan. Dengan menggunakan data latih terlabel, SVM akan mengidentifikasi pola tersembunyi dalam fitur dan melakukan klasifikasi untuk setiap berita baru. Dengan pendekatan ini, diharapkan SVM dapat memberikan keakuratan yang tinggi dalam membedakan antara berita *clickbait* dan *non-clickbait*, sehingga membantu mengurangi penyebaran konten *clickbait* yang menyesatkan dan meningkatkan kualitas informasi yang disajikan dalam berita *online*.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk membangun sebuah sistem yang dapat mendeteksi berita *clickbait*. Sistem ini diharapkan dapat secara efektif mengurangi penyebaran berita *clickbait* dan meningkatkan kualitas informasi yang disajikan dalam berita *online*.

2. METODE PENELITIAN

Tahapan-tahapan dari penelitian ini yang akan dijelaskan sebagai berikut:

1. *Requitmen analisis*
Pada proses pengumpulan data peneliti mendapatkan datanya dari penelitian CLIK.ID.
2. *Design*
Design atau perancangan sistem berperan penting karena disinilah dilakukannya representasi untuk sistem yang akan dibuat kedalam sebuah alur kerja sistem atau skema sistem sehingga dengan adanya perancangan sistem akan lebih terarah dalam proses pengembangan sistem nantinya.
3. *Development*
Tahapan ini yaitu mengimplentasikan metode yang dipilih ke dalam sistem memakai *source code*. Ini melibatkan penulisan, penyesuaian, dan pengujian kode program untuk memastikan efektivitas dalam konteks aplikasi.
4. *Testing*
Pengujian sistem dilaksanakan dengan maksud untuk memverifikasi apakah sistem yang dikembangkan telah beroperasi secara efektif dan sesuai dengan kebutuhan atau harapan dalam menjalankan fungsinya.
5. *Maintenance*
Tahap *maintenance* merupakan tahap akhir dari perancangan yang dibuat. *Developer* akan terus memperbaiki serta memelihara dari hasil evaluasi pada tahap sebelumnya.

2.1 Berita

berita adalah fakta dan gagasan terkini, akurat, menarik, dan dikenal luas merupakan bahan pembentuk berita. Publikasi termasuk surat kabar, radio, televisi, dan internet semuanya berperan dalam menyebarkan berita. Outlet berita hadir untuk mendidik masyarakat tentang peristiwa penting dan praktis [4].

2.2 Clickbait

Clickbait adalah judul atau tautan yang dirancang untuk menarik perhatian pembaca dengan cara yang provokatif atau menyesatkan, dengan tujuan utama untuk meningkatkan jumlah klik dan lalu lintas ke suatu situs web. Ciri-ciri berita clickbait ditandai dengan penggunaan judul yang emosional, berlebihan, atau menyesatkan. Judul-judul tersebut sering kali dimaksudkan untuk menarik perhatian pembaca, terlepas dari keakuratan atau kelengkapan informasi yang disampaikan [5].

2.3 Data Mining

Data mining adalah proses analisis data yang melibatkan penemuan pola-pola yang berharga atau berarti dalam kumpulan data yang besar, biasanya disebut sebagai database, untuk kemudian digunakan dalam pengambilan keputusan dan prediksi di masa depan. Pola-pola ini diidentifikasi dan dianalisis secara intuitif dengan menggunakan perangkat khusus [6].

2.4 Text Mining

Text mining adalah jenis data mining khusus yang bertujuan untuk menemukan pola menarik dalam sejumlah besar data berbentuk teks. Selain untuk klasifikasi, text mining juga berguna untuk menangani tugas-tugas seperti ekstraksi informasi (informasi ekstraksi), pengelompokan data (pengelompokan data), dan pencarian informasi (informasi pencarian) [7].

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode yang menghasilkan prediksi yang solid. Klasifikasi teks yang memanfaatkan bobot indeks kata sebagai fitur merupakan tujuan penggunaan metode SVM. Pada awal tahun 1992, Boser, Guyon, dan Vapnik membawa pembelajaran SVM ke khalayak yang lebih luas [8]. SVM merupakan suatu teknik yang mampu menyelesaikan permasalahan baik linier maupun non linier. Dalam ruang kerja berdimensi tinggi, ide kernel dipakai untuk menemukan *hyperplane* yang memaksimalkan margin antar kelas data, sehingga membantu mengatasi kesulitan non-linier. Kelas +1 dan kelas -1 merupakan dua kelompok dengan pola yang berbeda. *hyperplane* adalah alat yang hebat untuk membedakannya. Persamaan 1 sampai 2 merupakan persamaan dari *support vector machine linier*.

Persamaan 1 Menurut Vapnik dan Cortes (1995) dapat diperoleh:

$$X_i.W + b \geq 1 \text{ untuk } Y_i = 1 \quad (1)$$

$$X_i.W + b \geq -1 \text{ untuk } Y_i = -1 \quad (2)$$

Persamaan 2 mencari *hyperplane* digunakan untuk memaksimalkan jarak antara kelas data memakai persamaan berikut:

$$f(x) = w.x + b \quad (3)$$

Sedangkan ada beberapa fungsi non-linier seperti, linier, polinomial, radial basic funtion dan sigmoid . Dimana rumus sebagai berikut:

1. *Linier kernel*

$$K(x_p, x_d) = X_i^T X_j \quad (4)$$

2. *Polynomial*

$$K(x_p, x_d) = (X_i^T X_j + 1)^d, \gamma > 0 \quad (5)$$

3. *Radial Basic Funsion (RBF)*

$$K(x_p, x_d) = \exp(-\gamma \|X_i - X_j\|^2), \gamma > 0 \quad (6)$$

4. *Sigmoid*

$$K(x_p, x_d) = \tan \gamma X_i^T X_j + r \quad (7)$$

2.6 *Term Frecuency-Invers Document Frecuency (TF-IDF)*

Salah satu metode analisis teks yang digunakan untuk menentukan seberapa penting setiap kata dalam sebuah dokumen. Untuk melakukan ini, maka perlu menentukan frekuensi dokumen invers dan frekuensi term. Sementara TF menghitung jumlah kali kata muncul dalam sebuah dokumen, IDF menghitung berapa kali kata tersebut muncul secara unik di dalam korpus dokumen. Ketika kedua nilai ini digabungkan, TF-IDF menyorot kata-kata yang sering muncul dalam satu teks namun jarang muncul di dokumen lain selama pengambilan informasi dengan memberinya bobot lebih tinggi. Adapun beberapa tahapan dalam pembobotan *Term Frecuency-Inverse Document Frecuency (TF-IDF)* adalah:

1. Hitung *term frequency*(TF)

Term Frecuency ialah frekuensi kemunculan data dalam dokumen teks . *term frecuency*($tf_{t,d}$) didefenisikan bayaknya term t muncul dalam dokumen.

$$Wtf_{t,d} = \left\{ \begin{array}{l} 1 + \log 10 tf_{t,d}, \text{ if } tf_{t,d} > 0 \\ 0, \text{ otherwise} \end{array} \right\} \quad (8)$$

2. Hitung *bobot invers document frequency*(IDF)

Invers Dokument Frecuency yaitu hubungan kemunculan term diseluruh dokumen teks.

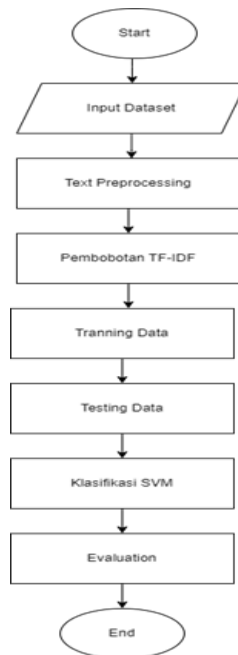
$$idf_t = \log \frac{N}{df_t} \quad (9)$$

3. Hitung *nilai bobot Term Frecuency-Invers Document Frecuency TF-IDF*

Nilai tf-idf dari suatu kata merupakan gabungan dari nilai tf dan nilai idf dalam perhitungan bobot.

$$W_{t,d} = Wtf_{t,d} \times idf_t \quad (10)$$

3. Skema Sistem



Gambar 1 skema sistem

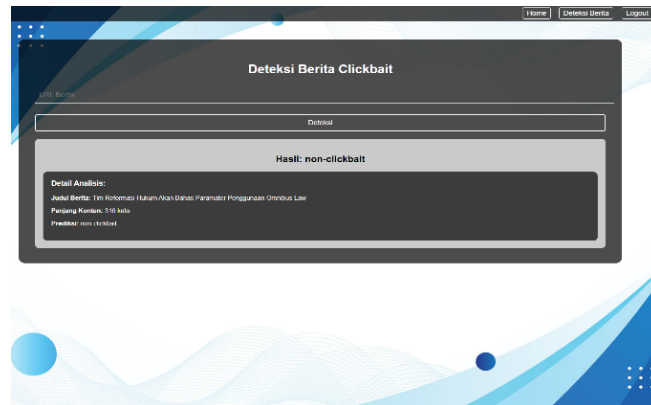
Sistem dimulai dengan mengidentifikasi masalah klasifikasi berita, yang melibatkan input dataset berita yang sudah dilabeli, dilanjutkan dengan langkah *text preprocessing* untuk membersihkan dan normalisasi teks, diikuti oleh pembobotan TF-IDF untuk mewakili teks dalam bentuk vektor. Langkah selanjutnya adalah training data dan testing data sehingga model klasifikasi SVM dapat dilatih pada separuh kumpulan data dan diuji pada separuh lainnya. Representasi vektor TF-IDF dipakai oleh SVM untuk klasifikasi kategori berita. Akurasi, presisi, perolehan, dan skor F1 adalah beberapa ukuran yang dipakai dalam evaluasi model. Pada akhirnya, sistem akan melakukan evaluasi model dengan cermat, dan model yang dilatih dapat dimanfaatkan dengan baik dengan mengategorikan berita baru.

4. HASIL DAN PEMBAHASAN

Penelitian ini juga mengembangkan sebuah sistem yang dapat mendeteksi berita clickbait dan non-clickbait dengan menerapkan metode *Support Vector Machine* (SVM). Metode SVM dipilih karena kemampuannya untuk mengklasifikasikan data secara efisien. Sistem ini bekerja dengan menerima input berupa URL dari berita yang ingin diperiksa, kemudian mengklasifikasikan berita tersebut ke dalam kategori *clickbait* atau *non-clickbait*. Model yang dikembangkan menggunakan SVM menunjukkan kinerja yang baik dengan akurasi yang memadai, sehingga sistem ini mampu memberikan hasil klasifikasi yang akurat dan cepat.

3.1 Hasil Implementasi Sistem

Tahapan implementasi sistem merupakan serangkaian langkah yang harus dilalui dalam proses pengembangan perangkat lunak yang telah dirancang, dianalisis, dan dibangun. Setelah proses pengembangan selesai, sistem akan melalui tahap pengujian untuk memastikan bahwa sistem tersebut memenuhi standar kelayakan dan berfungsi sesuai dengan spesifikasi serta tujuan yang telah ditetapkan. Hasil pengujian sistem dapat dilihat pada gambar dibawah ini.



Gambar 2 Pengujian sistem

Pada pengujian di atas, pengguna dapat dengan mudah memasukkan URL berita yang ingin dianalisis ke dalam sistem. Setelah URL dimasukkan, sistem secara otomatis mengambil judul dan konten dari artikel tersebut. Proses ini berlangsung cepat, sehingga pengguna tidak perlu melakukannya secara manual. Setelah data diambil, sistem akan menganalisis konten dan memberikan hasil klasifikasi yang jelas, menunjukkan apakah berita tersebut termasuk dalam kategori *clickbait* atau *non-clickbait*.

3.2 Pengumpulan Data

Penelitian ini menggunakan dataset berita berbahasa Indonesia yang diambil dari penelitian sebelumnya atau sumber CLIK.ID. data yang digunakan dalam penelitian ini yaitu judul dan konten berita yang telah diberi label berdasarkan kategori *clickbait* atau *non-clickbait*. Data yang digunakan dalam penelitian ini berjumlah 8.660 berita yang sudah dianotasi. data berita yang diperoleh dari CLICK.ID dapat dilihat pada gambar dibawah ini.

	judul	konten	label
0	Masuk Radar Pilwalkot Medan, Menantu Jokowi Be...	Medan - Menantu Jokowi, Bobby Afif Nasution, m...	non-clickbait
1	Malaysia Sudutkan RI: Isu Kabut Asap hingga In...	Jakarta - Malaysia kembali menyerang pemerinta...	non-clickbait
2	Viral! Driver Ojol di Bekasi Antar Pesanan Mak...	Bekasi - Seorang driver ojek online di Bekasi ...	clickbait
3	Kemensos Salurkan Rp 7,3 M bagi Korban Kerusuh...	Jakarta - Kementerian Sosial (Kemensos) menyal...	non-clickbait
4	MPR: Amandemen UUD 1945 Tak Akan Melebar ke Ma...	Jakarta - Anggota Badan Pengkajian MPR RI Bamb...	non-clickbait

Gambar 3 Data Berita

3.3 Preprocessing

Untuk membangun aplikasi deteksi berita *clickbait* dan *non-clickbait*, terdapat beberapa langkah penting yang dilakukan, terutama pada tahap *preprocessing*. *Preprocessing* pada *text mining* sangat penting dilakukan untuk meningkatkan kualitas data yang digunakan oleh model. Pada penelitian ini menggunakan teknik *preprocessing* yang dilakukan meliputi tokenisasi, di mana teks dipecah menjadi kata-kata atau token yang lebih kecil, penghilangan stopwords atau kata-kata yang tidak memiliki makna signifikan, serta stemming untuk mengubah kata-kata menjadi bentuk dasar. Berikut merupakan data latih berita yang sudah di tokenisasi, stopwords dan stemming dapat dilihat pada tabel dibawah ini.

Tabel 1 Tokenisasi, Stopword Dan Steaming

Preprocessing	Judul	konten	label
<i>Tokenisasi</i>	viral driver ojol di bekasi antar pesanan makanan pakai sepeda	seorang driver ojek online di bekasi eko susilo menjadi viral di media sosial karena mengantarkan pesanan makanan menggunakan sepeda	<i>Clickbait</i>
<i>Stopword</i>	viral driver ojol bekasi pesanan makanan pakai sepeda	seorang driver ojek online bekasi eko susilo viral media sosial mengantarkan pesanan makanan sepeda	<i>Clickbait</i>

Steaming viral driver ojol bekas pesan seorang driver ojek online bekas eko susilo
makan pakai sepeda viral media sosial antar pesan makan sepeda *Clickbait*

3.4 Term Frequency-Inverse Document Frequency TF-IDF

Pada penelitian ini, dilakukan pembobotan kata yang merupakan tahap penting dalam pemrosesan teks. Tahap pembobotan ini bertujuan untuk memberikan nilai bobot pada setiap kata dalam dokumen dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF adalah teknik yang menghitung seberapa sering suatu kata muncul dalam sebuah dokumen (*Term Frequency*) dan seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen (*Inverse Document Frequency*). Bobot yang dihasilkan oleh TF-IDF membantu dalam menentukan pentingnya sebuah kata dalam konteks tertentu, sehingga kata-kata yang lebih relevan akan memiliki bobot lebih tinggi, sementara kata-kata umum yang muncul di banyak dokumen akan diberi bobot yang lebih rendah. Dengan demikian, pembobotan kata menggunakan TF-IDF membantu meningkatkan akurasi model dalam mengenali pola pada teks, terutama untuk analisis seperti deteksi *clickbait*. Tabel dibawah ini merupakan Langkah dalam menghitung proses TF-IDF.

Tabel 2. Proses Perhitungan TF-IDF

TERM	TF	IDF	TF-IDF
viral	4	0,30103	1,20412
driver	3	0,30103	0,90309
ojol	3	0,30103	0,90309
bekas	5	0,30103	1,50515
pesan	4	0,30103	1,20412
makan	3	0,30103	0,90309
pakai	2	0,30103	0,60206
sepeda	5	0,30103	1,50515
seorang	1	0,30103	0,30103
ojek	1	0,30103	0,30103
online	1	0,30103	0,30103
eko	8	0,30103	0,30103
susilo	1	0,30103	2,40824
media	1	0,30103	0,30103
.....
sosial	1	0,30103	0,30103

3.5 Pembagian Data

Langkah ini bertujuan untuk membagi data menjadi data latih dan data uji. Pembagian ini bertujuan untuk melatih model deteksi *clickbait* menggunakan algoritma *machine learning* serta mengukur kinerja model berdasarkan hasil pengujian terhadap data uji.

Tabel 3. Pembagian Data

Keseluruhan Data	Data Latih	Data Uji
8.660	80% 6.888	20% 1.722

3.6 Pelatihan Model SVM

Implementasi *Support Vector Machine* (SVM) dalam penelitian ini menggunakan bahasa pemrograman python. Python memiliki *library Sklearn* yang dapat digunakan untuk mengimplementasikan SVM. Baris code implementasi SVM dapat dilihat pada gambar.

```

# Train SVM with linear kernel
model =SVC(C=1.0, kernel='linear', gamma='auto')
model.fit(X_train, y_train)

# Save model and vectorizer
joblib.dump(model, 'models/svm_clickbait_model.pkl')
joblib.dump(vectorizer, 'models/vectorizer.pkl')

# Make predictions on test data
y_pred = model.predict(X_test)

```

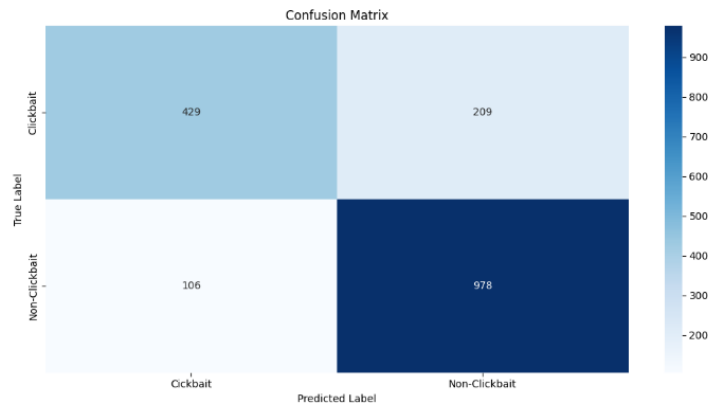
Gambar 4 Pemodelan SVM

Kode di atas merupakan *Model Support Vector Classifier (SVC)* dengan kernel linear digunakan untuk mengklasifikasikan berita antara clickbait dan non-clickbait. Dengan parameter $C=1.0$ dan $\text{gamma}='auto'$, model dilatih menggunakan fitur TF-IDF dari teks berita. Setelah pelatihan, model dan vectorizer disimpan dalam file untuk penggunaan ulang. Model diuji dengan data uji, dan hasil evaluasi dilakukan menggunakan confusion matrix untuk menilai performa prediksi.

3.7 Evaluasion

Evaluasi ini dilakukan untuk mengetahui tingkat keberhasilan kinerja model dalam melakukan proses klasifikasi *Clickbait* dan *Non-Clickbait*. dari *confusion matrix* dapat diketahui akurasi, presisi dan recall. Pada penelitian ini menggunakan 2 kelas sehingga parameter perhitungan *confusion matrix* terdiri dari 4 hal yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*.

Data yang digunakan dalam pelatihan terdiri dari 6.888 berita, sedangkan data pengujian berjumlah 1.772 berita. Berdasarkan tabel *confusion matrix*, hasil prediksi menunjukkan bahwa:



Gambar 5 Confusion matrix

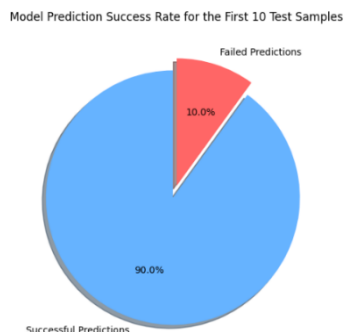
Dari *confusion matrix* diatas kita bisa menghitung beberapa metrik evaluasi penting:

Classification Report:				
	precision	recall	f1-score	support
clickbait	0.80	0.67	0.73	638
non-clickbait	0.82	0.90	0.86	1084
accuracy			0.82	1722
macro avg	0.81	0.79	0.80	1722
weighted avg	0.82	0.82	0.81	1722

Gambar 6 Klasifikasi report

Berdasarkan hasil pengujian mendeteksi *clickbait* dengan menggunakan kernel linier mendapatkan akurasi sebesar *accuracy* 82%, *precision* 81%, *recall* 65%, dan *F1-score* 72% di atas dengan metode SVM dalam

3.8



Grafik Keberhasilan Model

Gambar 7 Tingkat Keberhasilan model

Grafik ini menunjukkan bahwa model memiliki tingkat keberhasilan sebesar 90% dan tingkat kegagalan hanya 10% untuk 10 data uji pertama. Ini mengindikasikan bahwa model mampu memprediksi dengan baik antara *clickbait* dan *non-clickbait*.

4. DISKUSI

Hasil penelitian ini menunjukkan bahwa metode *Support Vector Machine* (SVM) yang diterapkan untuk mendeteksi berita *clickbait* dan *non-clickbait* mampu memberikan kinerja yang baik. Model ini mencapai akurasi sebesar 82%, *precision* 81%, *recall* 65%, dan *F1-score* 72%. Selain itu, grafik yang dihasilkan memperlihatkan tingkat keberhasilan model mencapai 90%, dengan tingkat kegagalan hanya 10% untuk 10 data uji pertama. Ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam membedakan antara *clickbait* dan *non-clickbait*. Kinerja ini menggarisbawahi efektivitas SVM dalam mengklasifikasikan teks berita berdasarkan karakteristik *clickbait*.

Penelitian yang dilakukan oleh Wanda Athira Luqyana berjudul Analisis Sentimen Cyberbullying pada Komentar Instagram dengan metode klasifikasi Support Vector Machine [3], menunjukkan hasil klasifikasi yang membedakan antara kelas positif dan negatif, dengan fokus pada sentimen positif dan negatif terkait cyberbullying. Berdasarkan pengujian, tingkat akurasi tertinggi yang dicapai adalah 90%, disertai *precision* 94,44%, *recall* 85%, dan *F-measure* 89,47%. Penelitian ini menggunakan komposisi data latih 50% dan data uji 50%.

Jika dibandingkan, penelitian ini mengenai klasifikasi berita *clickbait* menggunakan model *Support Vector Classifier* (SVC) memperoleh akurasi sebesar 82%, *precision* 81%, *recall* 65%, dan *F1-score* 72%. Meskipun hasil akurasi dan *precision* menunjukkan performa yang baik, nilai *recall* yang lebih rendah mengindikasikan bahwa model belum sepenuhnya optimal dalam mendeteksi semua berita *clickbait*. Hal ini mungkin disebabkan oleh ketidakseimbangan data antara kelas *clickbait* dan *non-clickbait*, di mana jumlah berita *clickbait* jauh lebih sedikit.

5. KESIMPULAN

Dari hasil pengujian yang telah dilakukan, dapat ditarik beberapa kesimpulan dalam mendeteksi berita *clickbait* pada yang menerapkan metode SVM. Hal yang disimpulkan dari penelitian ini adalah sebagai berikut:

1. Penelitian ini mengembangkan sistem yang dapat deteksi berita *clickbait* dan *non-clickbait* dengan menerapkan metode *Support Vector Machine* (SVM). Sistem ini berfungsi dengan menerima input berupa URL dari berita yang ingin diperiksa, lalu mengklasifikasikan berita tersebut kedalam kategori *clickbait* dan *non-clickbait*. Model yang dikembangkan menggunakan metode SVM menunjukkan kinerja yang baik dengan akurasi yang memadai.
2. Model *Support Vector Classifier* (SVC) dengan kernel linear digunakan untuk mengklasifikasikan berita antara *clickbait* dan *non-clickbait*. Hasil pengujian menunjukkan akurasi sebesar 82%, *precision* 81%, *recall* 65%, dan *F1-score* 72%. Selain itu, grafik menunjukkan bahwa model memiliki tingkat keberhasilan sebesar

90% dan tingkat kegagalan hanya 10% untuk 10 data uji pertama. Ini mengindikasikan bahwa model mampu memprediksi dengan baik antara *clickbait* dan *non-clickbait*.

6. UCAPAN TERIMA KASIH

Terima kasih yang sebesar-besarnya saya sampaikan kepada dosen pembimbing yang telah memberikan bimbingan, serta kepada orang tua dan keluarga yang telah memberikan dukungan dan bantuan tanpa henti selama proses penelitian ini. Dukungan moril dan materiil yang diberikan sangat berarti, sehingga penelitian ini dapat berjalan dengan lancar dan mencapai hasil yang lebih baik. Ucapan terima kasih juga saya sampaikan kepada Universitas Malikussaleh yang telah memberikan fasilitas untuk menunjang penelitian ini. Semoga segala kebaikan dan bantuan yang diberikan mendapatkan balasan yang setimpal.

7. DAFTAR PUSTAKA

- [1] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Adriansyah, M. H. Kafabi, and S. Sudioanto, "Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 85–89, 2022, doi: 10.30591/jpit.v7i2.3401.
- [2] A. F. Yavi, "Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [3] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] E. P. Bangun, F. V. I. A Koagouw, and J. S. Kalangi, "Analisis Isi Unsur Kelengkapan Berita Pada Media Online Manadopostonline.com," *Acta Diurna Komun.*, vol. 1, no. 3, pp. 4–13, 2019, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/actadiurnakomunikasi/article/view/25560>
- [5] R. Sagita, U. Enri, and A. Primajaya, "Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN)," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 2, pp. 230–239, 2020, doi: 10.33633/joins.v5i2.3705.
- [6] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [7] R. Tjut, A. Adek, H. Al Kautsar Aidilof, and W. Kurniawan, "Klasifikasi Berita Olahraga Pada Portal Berita Online Dengan Metode K-Nearest Neighbour (Knn) Dan Levenshtein Distance," *J. Teknol. Terap. Sains 4.0*, vol. 2, no. 1, p. 365, 2021, doi: 10.29103/tts.v2i1.3760.
- [8] I. P. Monika and M. T. Furqon, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3165–3166, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>

SENASTIKA Universitas Malikussaleh

DETEKSI KANKER KULIT MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK

Cantika Serenita¹, Fatih Dwi Laksana², Rizka Salsabila Nasution³, Munirul Ula⁴.

^{1,2,3,4}Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email : ¹cantika.210170056@mhs.unimal.ac.id , ²fatih.210170057@mhs.unimal.ac.id , ³rizka.210170054@mhs.unimal.ac.id , ⁴munirulula@unimal.ac.id .

Abstrak

Kanker kulit merupakan penyakit berbahaya yang terjadi akibat perubahan abnormal sel kulit, yang dapat menyebabkan penyebaran kanker ke organ lain. Deteksi dini kanker kulit sangat penting untuk meningkatkan peluang kesembuhan, namun metode konvensional seperti biopsi memerlukan waktu dan biaya yang besar. Dalam beberapa tahun terakhir, teknik berbasis deep learning seperti Convolutional Neural Network (CNN) telah menunjukkan potensi besar dalam deteksi otomatis penyakit berbasis gambar medis. Penelitian ini bertujuan untuk membandingkan performa dua arsitektur CNN, yaitu VGG16 dan ResNet50, dalam klasifikasi kanker kulit menggunakan dataset dari Kaggle. Dataset terdiri dari 2.967 gambar yang dibagi menjadi dua kategori: malignant dan benign. Proses penelitian meliputi pengumpulan data, preprocessing, pelatihan model, dan evaluasi menggunakan metrik seperti akurasi, precision, recall, dan f1-score. Hasil penelitian menunjukkan bahwa model VGG16 mencapai akurasi sebesar 88,79%, sementara ResNet50 mencapai akurasi 86,97%. Meskipun VGG16 menunjukkan hasil yang lebih baik dalam akurasi, model ini rentan terhadap overfitting. Penelitian ini memberikan kontribusi pada pengembangan sistem deteksi kanker kulit otomatis yang lebih cepat dan efisien, serta diharapkan dapat mendukung peningkatan diagnosis dini dan pengobatan kanker kulit.

Keywords : *Convolutional Neural Network (CNN), Kanker kulit, Klasifikasi gambar medis, ResNet50, VGG16*

1. PENDAHULUAN

Kanker kulit merupakan penyakit yang disebabkan oleh perubahan karakteristik sel penyusun kulit dari normal menjadi ganas. Hal ini menyebabkan sel membelah secara tidak terkendali dan merusak DNA, yang pada akhirnya merusak jaringan kulit serta meningkatkan risiko penyebaran kanker ke organ tubuh lainnya [1]. Mengingat kulit adalah organ yang terpapar langsung dengan sinar matahari, paparan berlebihan terhadap sinar UV yang berbahaya dapat memicu perkembangan kanker kulit. Penipisan lapisan ozon akibat perubahan lingkungan global, seperti efek global warming, telah menyebabkan peningkatan radiasi matahari, sehingga risiko kanker kulit juga semakin meningkat [2].

Di Indonesia, kanker kulit menempati urutan ketiga setelah kanker rahim dan kanker payudara, dengan prevalensi sebesar 5,9-7,8% dari semua jenis kanker setiap tahunnya. Beberapa faktor risiko utama adalah kulit putih, paparan matahari berlebihan, dan riwayat keluarga dengan kanker kulit [3]. Meskipun proses diagnosis umumnya dilakukan melalui biopsi dan pemeriksaan mikroskopis, teknik ini memerlukan waktu dan biaya yang signifikan [2]. Jika kanker kulit terlambat terdeteksi, tingkat kelangsungan hidup pasien hanya sekitar 14%, sedangkan deteksi dini dapat meningkatkan peluang hidup hingga lebih dari 99%[1]. Oleh karena itu, deteksi dini sangat penting untuk meningkatkan efektivitas pengobatan kanker kulit [4].

Berbagai teknik klasifikasi berbasis citra telah dikembangkan untuk mendukung deteksi kanker kulit, termasuk metode deep learning yang memanfaatkan Convolutional Neural Network (CNN) [5]. CNN telah menunjukkan kemajuan signifikan dalam pengenalan gambar dan digunakan dalam deteksi dini penyakit berbasis citra medis [6]. CNN dapat mengotomatisasi klasifikasi gambar tanpa memerlukan ekstraksi fitur manual, menjadikannya pilihan yang cepat dan akurat untuk klasifikasi kanker kulit[7]. Model VGG16 dan ResNet50 merupakan dua arsitektur CNN yang sering digunakan dan menunjukkan performa baik dalam pengenalan gambar

medis [8]. Pemilihan arsitektur VGG dan ResNet-50 didasarkan pada kemampuan ekstraksi fitur yang unggul, kinerja yang telah terbukti mutakhir, serta kesesuaiannya untuk penerapan transfer learning[9].

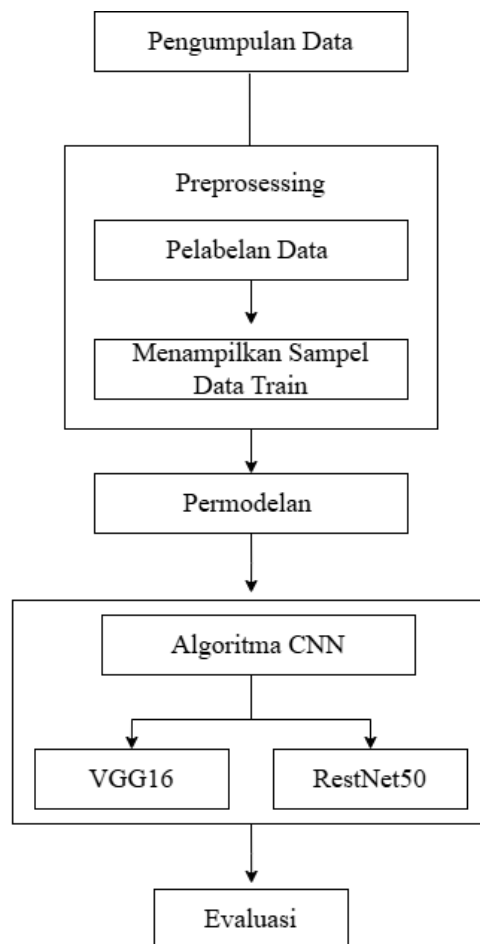
VGG16 adalah model dengan 16 lapisan konvolusi yang efektif untuk klasifikasi gambar [10]. ResNet50 menggunakan pendekatan skip connections yang memungkinkan jaringan belajar dari fitur yang lebih dalam tanpa mengalami degradasi performa [11]. RestNet adalah salah satu model yang paling mengesankan dan tidak rumit untuk tugas pengenalan gambar[12]. Dalam penelitian sebelumnya, ResNet50 dengan skema augmentasi data telah mencapai akurasi sebesar 83%. Meski demikian, untuk mencapai akurasi lebih tinggi hingga 90%, diperlukan peningkatan ukuran dataset melalui penggunaan dataset seperti ImageNet[13]. Penelitian ini akan mengeksplorasi perbandingan performa kedua arsitektur ini dalam mendukung klasifikasi kanker kulit.

Penelitian ini bertujuan untuk membandingkan performa arsitektur VGG16 dan ResNet50 dalam klasifikasi kanker kulit menggunakan dataset ImageNet. Dengan menggunakan pendekatan deep learning berbasis CNN, penelitian ini diharapkan dapat menentukan arsitektur yang paling optimal dalam menghasilkan klasifikasi kanker kulit yang akurat dan efisien.

Penelitian ini penting karena memberikan kontribusi terhadap pengembangan teknologi diagnosa otomatis berbasis AI yang lebih cepat dan akurat untuk deteksi dini kanker kulit. Hasil dari penelitian ini diharapkan dapat membantu dalam mempercepat proses diagnosis dan pengobatan, serta mengurangi angka kematian akibat keterlambatan deteksi kanker. Selain itu, penelitian ini juga dapat menjadi acuan bagi pengembangan lebih lanjut aplikasi klinis yang menggunakan arsitektur CNN untuk mendukung deteksi dan klasifikasi penyakit berbasis gambar medis.

2. METODE PENELITIAN

Penelitian ini dilakukan untuk membandingkan algoritma CNN dengan mengevaluasi hasil akurasi dari arsitektur VGG16 dan RestNet50. Alur penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Penelitian ini memanfaatkan data yang diperoleh dari situs Kaggle.com, yang merupakan salah satu platform populer untuk berbagi dataset. Data tersebut diambil dari halaman yang berfokus pada Skin Cancer ISIC, yang merupakan sumber dataset terkait kanker kulit. Dataset ini terdiri dari total 2.967 gambar, di mana gambar-gambar tersebut menggambarkan berbagai jenis kanker kulit dan telah dikelompokkan ke dalam 2 kategori atau kelas yang berbeda. Selanjutnya, data ini diproses secara mendalam dengan tujuan untuk melakukan klasifikasi jenis-jenis kanker kulit, sehingga dapat membantu dalam identifikasi dan diagnosis lebih lanjut. Melalui pengolahan data yang cermat, penelitian ini bertujuan untuk memberikan kontribusi dalam bidang kesehatan, khususnya dalam mendukung teknologi pengenalan gambar medis untuk mendeteksi kanker kulit secara otomatis.

2.2 Pre-processing

Tahap preprocessing sangat penting untuk memastikan data siap diproses oleh model, Tahapan selanjutnya adalah preprocessing, yaitu resizing, yang dilakukan untuk menyeragamkan ukuran citra[14]. Dua langkah yang krusial dalam preprocessing adalah pelabelan data dan menampilkan sampel data train. Pelabelan data adalah proses di mana setiap gambar dalam dataset dikategorikan sesuai dengan kelas atau jenis kanker kulit yang diidentifikasi. Dalam penelitian ini, dataset kanker kulit biasanya terdiri dari gambar yang mewakili beberapa jenis kanker kulit (misalnya, melanoma, basal cell carcinoma, actinic keratosis, dll.) dan gambar yang sehat. Setiap gambar diberikan label sesuai dengan kelasnya. Setelah data dilabeli, langkah penting selanjutnya dalam preprocessing adalah menampilkan sampel gambar dari dataset train. Langkah ini berguna untuk memeriksa apakah data telah dimuat dengan benar dan apakah label yang terkait dengan setiap gambar telah ditetapkan secara akurat.

2.3 Permodelan

Setelah proses preprocessing data selesai dilakukan, tahap selanjutnya adalah melatih model Convolutional Neural Network (CNN). Penilaian kinerja model akan menunjukkan kemampuannya untuk secara akurat mengidentifikasi penyakit secara keseluruhan[15]. Penelitian ini memanfaatkan teknik CNN dengan menerapkan dua arsitektur jaringan yang populer dan berbeda untuk mendapatkan hasil yang optimal, yaitu VGG16 dan ResNet50. VGG16 merupakan arsitektur yang dikenal sederhana namun efektif, dengan kedalaman 16 lapisan, yang dirancang untuk menghasilkan representasi fitur visual yang kaya. Sementara itu, ResNet50 adalah arsitektur yang lebih kompleks dengan 50 lapisan, yang mengatasi masalah vanishing gradient dan memungkinkan pelatihan jaringan lebih dalam dengan cara memanfaatkan residual connections.

2.4 Evaluasi

Pada tahap ini, model dievaluasi menggunakan data uji. Confusion matrix digunakan sebagai alat untuk mengukur kinerja klasifikasi kanker kulit, termasuk akurasi, presisi, recall, dan f1-score. Hasil pengujian disajikan dalam bentuk tabel yang terstruktur sebagai confusion matrix. Perhitungan akurasi model dilakukan dengan membagi jumlah data yang diklasifikasikan dengan benar oleh model dengan total data yang tersedia. Proses ini dapat dijelaskan melalui persamaan tertentu, yang juga dapat digunakan untuk menghitung akurasi, presisi, dan recall berdasarkan confusion matrix.

$$Akurasi = (TP+TN) / (TP+FP+FN+TN)$$

$$Recall = (TP) / (TP+FN)$$

$$Precision = (TP) / (TP+FP)$$

Keterangan:

TP (True Positif) = Jumlah data yang diklasifikasikan dengan benar sebagai positif dan sesuai dengan kelas yang dimaksud.

TN (True Negatif) = Jumlah data yang diklasifikasikan dengan benar sebagai negative untuk kelas yang dimaksud (tidak termasuk TP kelas itu).

FP (False Positif) = Jumlah data yang salah diklasifikasikan sebagai positif untuk kelas yang dimaksud (tidak termasuk TP kelas itu).

FN (False Negatif) = Jumlah data yang salah diklasifikasikan sebagai negative untuk kelas yang dimaksud (tidak termasuk dari baris dan kolom dari kelas tersebut).

3. HASIL DAN PEMBAHASAN

Dataset yang dipakai dalam penelitian ini adalah kumpulan data gambar yang diambil pada website Kaggle.

3.1 Pengumpulan Data

Tabel 1 Jenis Kanker yang di Teliti

No	Jenis Kanker Kulit	Jumlah Citra
1.	Malignant (Ganas)	1331
2.	Benign (Jinak)	1636

3.2

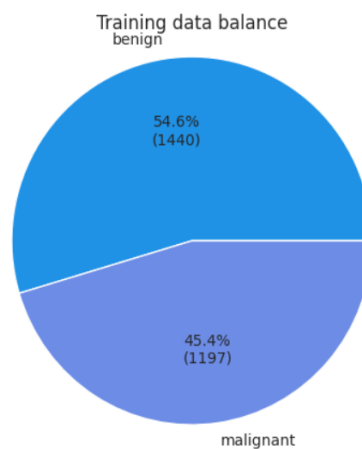
Tahap adalah langkah melanjutkan ke gambar. Pada tahap awal ini, dilakukan inisialisasi sebagaimana dijelaskan dalam Tabel 2.

Preprocessing preprocessing pertama sebelum pemrosesan

Tabel 2 Inisialisasi Hyperparameter

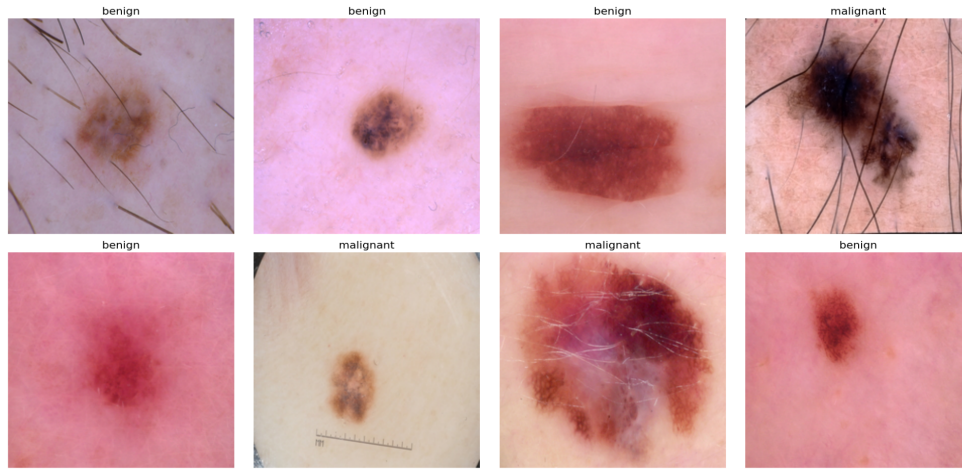
Model	Epoch	Learning Rate	Batch Size	Input Size
VGG16	15	0.0001	8	224 x 224 x 3
RestNet50	15	0.0001	8	224 x 224 x 3

Setelah proses inisialisasi selesai, langkah berikutnya adalah melakukan pemeriksaan terhadap data pelatihan untuk mengetahui apakah data tersebut seimbang atau tidak, guna menghindari bias pada data.



Gambar 2. Diagram Keseimbangan Data Pelatihan

Setelah selesai melakukan pemeriksaan terhadap data pelatihan, selanjutnya data training ditampilkan dalam bentuk gambar berikut ini :



Gambar 3. Contoh Gambar Data Pelatihan

3.3 Permodelan

Dalam penelitian ini terdapat 2.967 gambar kanker kulit yang terbagi menjadi dua bagian. Sebanyak 11% atau 328 gambar digunakan untuk data testing dan sebanyak 89% atau 2.637 gambar untuk data training. Setelah preprocessing dataset selesai, langkah berikutnya adalah merancang model yang akan digunakan. Pada penelitian ini, digunakan dua arsitektur model CNN, yaitu VGG16 dan ResNet50.

A. Perancangan model RestNet50

conv5_block2_add (Add)	(None, 7, 7, 2048)	0	conv5_block1_out[0][0... conv5_block2_3_bn[0][...
conv5_block2_out (Activation)	(None, 7, 7, 2048)	0	conv5_block2_add[0][0]
conv5_block3_1_conv (Conv2D)	(None, 7, 7, 512)	1,049,088	conv5_block2_out[0][0]
conv5_block3_1_bn (BatchNormalization)	(None, 7, 7, 512)	2,048	conv5_block3_1_conv[0...
conv5_block3_1_relu (Activation)	(None, 7, 7, 512)	0	conv5_block3_1_bn[0][...
conv5_block3_2_conv (Conv2D)	(None, 7, 7, 512)	2,359,808	conv5_block3_1_relu[0...
conv5_block3_2_bn (BatchNormalization)	(None, 7, 7, 512)	2,048	conv5_block3_2_conv[0...
conv5_block3_2_relu (Activation)	(None, 7, 7, 512)	0	conv5_block3_2_bn[0][...
conv5_block3_3_conv (Conv2D)	(None, 7, 7, 2048)	1,050,624	conv5_block3_2_relu[0...
conv5_block3_3_bn (BatchNormalization)	(None, 7, 7, 2048)	8,192	conv5_block3_3_conv[0...
conv5_block3_add (Add)	(None, 7, 7, 2048)	0	conv5_block2_out[0][0... conv5_block3_3_bn[0][...
conv5_block3_out (Activation)	(None, 7, 7, 2048)	0	conv5_block3_add[0][0]
global_average_pooling2d (GlobalAveragePooling2D)	(None, 2048)	0	conv5_block3_out[0][0]
batch_normalization (BatchNormalization)	(None, 2048)	8,192	global_average_poolin...
dense (Dense)	(None, 128)	262,272	batch_normalization[0...
dropout (Dropout)	(None, 128)	0	dense[0][0]
dense_1 (Dense)	(None, 32)	4,128	dropout[0][0]
dense_2 (Dense)	(None, 2)	66	dense_1[0][0]

Total params: 23,862,370 (91.03 MB)
 Trainable params: 4,736,226 (18.07 MB)
 Non-trainable params: 19,126,144 (72.96 MB)

Gambar 4. Perancangan Model RestNet50

Perancangan model ResNet50 dapat dilihat pada Gambar 4. model ini terdiri dari 181 layer. Yang mana ini termasuk dari semua jenis layer, mulai dari layer convolutional, batch normalization, activation, dan lain-lain. untuk parameter layer yang dapat dilatih dan tidak dapat dilatih berjumlah 23.862.370, dimana parameter yang dapat dilatih sebanyak 4.736.226 dan parameter yang tidak dapat dilatih terdapat 19.126.144 . pada layer awal model, terdapat inputan gambar dengan ukuran 224 x 224 x 3 (RGB image) tanpa parameter yang dilatih. secara keseluruhan layer awal dari arsitektur ResNet50 ini diikuti oleh banyak residual block yang dapat membantu menangani degradasi performa seiring bertambahnya kedalaman. fully connected (Dense) layer menghubungkan vektor hasil pooling ke output layer. dan untuk output akhir dari layer ini terdapat pada layer Dense_2 yang berukuran (none, 2), dimana “none” menunjukkan batch size yang bervariasi, dan “2” menunjukkan dua kelas output.

B. Perancangan Model VGG16

block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_max_pooling2d (GlobalMaxPooling2D)	(None, 512)	0
flatten (Flatten)	(None, 512)	0
dense_3 (Dense)	(None, 512)	262,656
dropout_1 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 256)	131,328
dense_5 (Dense)	(None, 128)	32,896
dense_6 (Dense)	(None, 32)	4,128
dense_7 (Dense)	(None, 2)	66

Total params: 15,145,762 (57.78 MB)
Trainable params: 7,510,498 (28.65 MB)
Non-trainable params: 7,635,264 (29.13 MB)

Gambar 5. Perancangan Model VGG16

Pada Gambar 5 merupakan perancangan model VGG16. Model ini terdiri dari 27 layer. Layer tersebut terdiri dari Convolutional Layers, Pooling Layers, Global Max pooling, Flatten Layer, Dense Layers dan Dropout Layer. Total jumlah keseluruhan parameter layer sebesar 15.145.762, dimana jumlah parameter terlatih sebanyak 7.510.498 dan parameter tidak terlatih sebesar 7.635.264 layers. Lapisan input (InputLayer) dalam model berfungsi untuk menerima data berupa gambar dengan dimensi 224x224 piksel dan 3 saluran warna (RGB). Bentuk output lapisan ini ditunjukkan sebagai (None, 224, 224, 3), di mana None menunjukkan ukuran batch yang dinamis. Lapisan ini tidak memiliki parameter yang dilatih, karena hanya berperan sebagai penerima data awal. Output akhir dari layer ini terdapat pada layer Dense_7 yang berukuran (none, 2), dimana “none” menunjukkan batch size yang bervariasi, dan “2” menunjukkan dua kelas output.

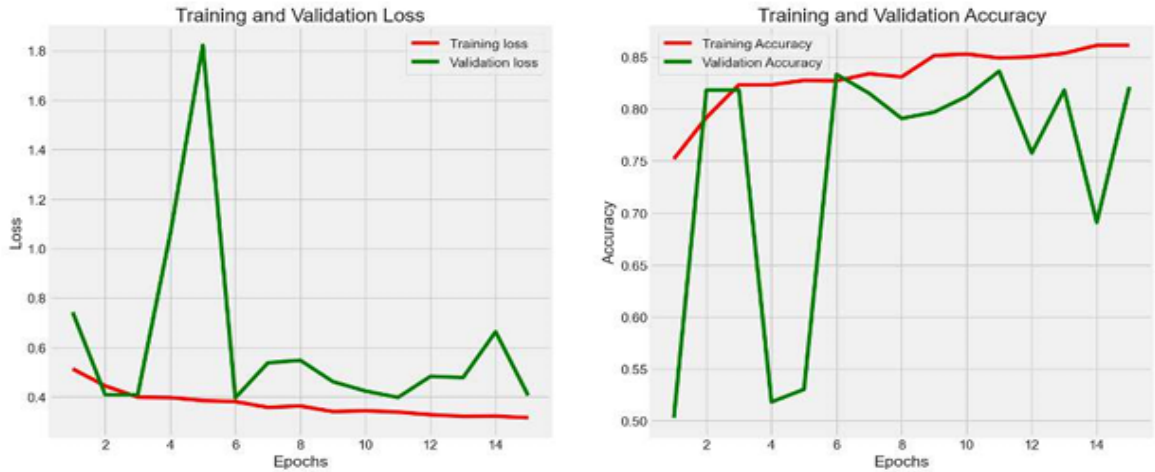
3.4 Evaluation

Pada tahap ini, evaluasi model dilakukan dengan melihat hasil dari nilai accuracy, precision, dan recall. Pengujian tingkat accuracy model dilakukan menggunakan confusion matrix untuk mengukur seberapa baik model

dalam mengklasifikasikan jenis kanker kulit. Semakin tinggi nilai accuracy yang diperoleh, semakin baik kualitas model yang dikembangkan.

A. Accuracy

Model VGG16 dan RESNet50 menampilkan grafik akurasi data train dan data test pada gambar dibawah ini. dimana, digunakan jumlah iterasi sebanyak 15 Epoch untuk melatih model.



Gambar 6. Kurva Accuracy Model RestNet50

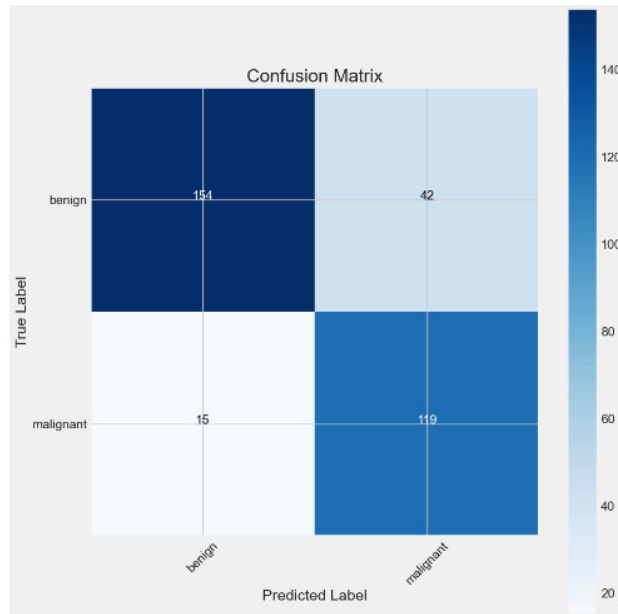
Gambar 6. Model RestNet berhasil mencapai akurasi sebesar 87% pada training accuracy, sedangkan validation accuracy mencapai nilai 82%



Gambar 7. Kurva Accuracy Model VGG16

Gambar 7. VGG16 Sedangkan model VGG16 berhasil mencapai akurasi sebesar 99% pada training accuracy, dan mencapai akurasi 86% pada validation accuracy.

B. Confusion Matrix



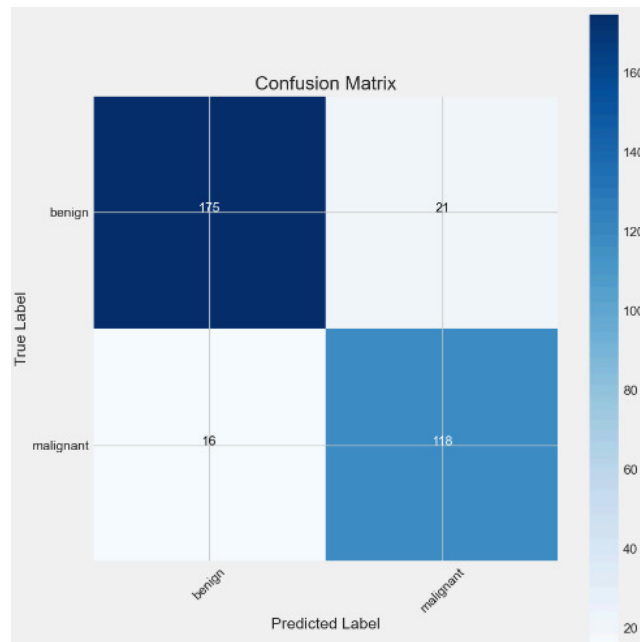
Gambar 8. Confusion Matrix RestNet50

Tabel 3 merupakan pemaparan performa kinerja model RestNet50 yang mencakup precision, recall dan f1-score.

Tabel 3. Hasil Evaluasi Kinerja RestNet50

Kelas	Precision	Recall	F1-Score
Malignan	73.91%	88.81%	80.68%
Bening	91.12%	78.57%	84.38%

Berikut Confusion Matrix Untuk Matrix VGG16 :



Gambar 9. Confusion Matrix VGG16

Tabel 4 Merupakan pemaparan performa kinerja model VGG16 yang mencakup Precision, Recall, dan f1-score.

Tabel 4. Hasil Evaluasi Kinerja VGG16

Kelas	Precision	Recall	F1-Score
Malignan	84.89%	88.06%	86.45%
Bening	91.62%	89.29%	90.44%

Setelah menyelesaikan tahapan permodelan untuk kedua arsitektur dengan membagi data menjadi 11% untuk data testing dan 89% sebagai data training, Tabel 5 menunjukkan hasil perbandingan antara RestNet50 dan VGG16 dalam penelitian ini.

Tabel 5. Hasil Komparasi Model RestNet50 dengan VGG16

Model	Akurasi	Presisi	Recall	F1-Score
RESNet50	86.97%	83.21%	85.07%	84.13%
VGG16	88.79%	84.89%	88.06%	86.45%

dari hasil penelitian ini terdapat tingkat akurasi pada model VGG16 dan RESNet50 yang mana, bisa dibandingkan dengan beberapa literatur yang menjadi acuan dalam pembuatan jurnal ini.

4. DISKUSI

Penelitian ini membandingkan performa dua arsitektur Convolutional Neural Network (CNN), yaitu VGG16 dan ResNet50, dalam klasifikasi kanker kulit. Hasil penelitian menunjukkan bahwa VGG16 mencapai akurasi sebesar 88.79%, lebih unggul dibandingkan dengan ResNet50 yang mencapai akurasi 86.97%. VGG16 juga menghasilkan nilai precision, recall, dan F1-score yang lebih baik untuk kedua kelas (Malignant dan Benign). Pada kelas Malignant, VGG16 memperoleh precision sebesar 84.89%, recall 88.06%, dan F1-score 86.45%, sedangkan ResNet50 hanya menghasilkan precision 83.21%, recall 85.07%, dan F1-score 84.13%.. model dalam penelitian ini bisa dibandingkan dengan beberapa literatur yang menjadi acuan dalam pembuatan jurnal.

Jika dibandingkan dengan penelitian lain, yang juga menggunakan model CNN untuk mendeteksi kanker kulit memberikan hasil akurasi sebesar 99,70% dengan arsitektur VGG16. Dan pada referensi lain, didapatkan hasil akurasi untuk arsitektur RESNet50 adalah sebesar 94,88%. Hal ini dipengaruhi oleh penggunaan dataset yang lebih besar, teknik augmentasi data yang lebih ekstensif, atau tuning hyperparameter yang lebih optimal. Tentunya semakin banyak dataset, maka model akan belajar lebih banyak untuk mengembangkan data tersebut, sehingga bisa memberikan hasil yang lebih bagus.

Meskipun demikian, hasil penelitian ini menunjukkan bahwa kedua arsitektur memiliki potensi yang baik untuk klasifikasi kanker kulit, dengan VGG16 sedikit lebih unggul. Untuk meningkatkan performa model di masa depan, diperlukan strategi tambahan seperti augmentasi data yang lebih baik, penggunaan transfer learning dengan dataset yang lebih besar, serta tuning hyperparameter yang lebih teliti.

5. KESIMPULAN

Dalam penelitian ini, diketahui bahwa model yang digunakan untuk melakukan perbandingan adalah model arsitektur VGG16 dan RESNet50. Dengan jumlah dataset sebanyak 2967 citra kanker kulit yang dibagi menjadi data train 2637 gambar dan data tes sebanyak 330 gambar. Dari kedua model tersebut, VGG16 mendapatkan akurasi tertinggi dengan akurasi data train sebesar 99% dan data test 88%. sedangkan untuk model RESNet50, data train nya memiliki akurasi 87% dan data test 82%.

DAFTAR PUSTAKA

- [1] R. AGUSTINA, R. MAGDALENA, and N. K. C. PRATIWI, "Klasifikasi Kanker Kulit menggunakan Metode Convolutional Neural Network dengan Arsitektur VGG-16," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 2, p. 446, Apr. 2022, doi: 10.26760/elkomika.v10i2.446.

- [2] D. A. Nurlitasari, R. Magdalena, and R. Y. N. Fu'adah, "ANALISIS PERFORMANSI SISTEM KLASIFIKASI KANKER KULIT MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK," *JOURNAL OF ELECTRICAL AND SYSTEM CONTROL ENGINEERING*, vol. 5, no. 2, pp. 91–99, Feb. 2022, doi: 10.31289/jesce.v5i2.5691.
- [3] J. Setiabudi, M. Wardhana, I. G. A. A. E. Indira, and N. M. D. Puspawati, "PROFIL PRA KANKER DAN KANKER KULIT DI RSUP SANGLAH PERIODE 2015-2018," *JURNAL MEDIKA UDAYANA*, vol. 10, no. 3, pp. 83–88, 2021, doi: <https://doi.org/10.24843/mu.2021.v10.i3.p13>.
- [4] Orlando and M. Ezar Al Rivian, "KLASIFIKASI JENIS KANKER KULIT MANUSIA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK," *jurnal.mdp.ac.id*, vol. 2, no. 1, pp. 144–150, 2023, doi: <https://doi.org/10.35957/mdp-sc.v2i1.4335>.
- [5] D. Septhya, I. Engineering, and S. Amik Riau, "Penerapan Algoritma Convolutional Neural Network Untuk Klasifikasi Penyakit Kanker Kulit," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, pp. 6590–6600, 2024, doi: 10.33022/ijcs.v13i4.4262.
- [6] Q. Aina Fitroh and S. 'Uyun, "Deep Transfer Learning untuk Meningkatkan Akurasi Klasifikasi pada Citra Dermoskopi Kanker Kulit," *JURNAL NASIONAL TEKNIK ELEKTRO DAN TEKNOLOGI INFORMASI*, vol. 12, no. 2, pp. 78–84, 2023, doi: <https://doi.org/10.22146/jnteti.v12i2.6502>.
- [7] A. A. Yonismara and A. Salam, "Implementasi Model Convolutional Neural Network (CNN) pada Aplikasi Deteksi Kanker Kulit Menggunakan Expo React Native," *Technology and Science (BITS)*, vol. 6, no. 1, pp. 226–235, 2024, doi: 10.47065/bits.v6i1.5351.
- [8] A. Maharil, "PERBANDINGAN ARSITEKTUR VGG16 DAN RESNET50 UNTUK REKOGNISI TULISAN TANGAN AKSARA LAMPUNG," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, pp. 236–243, 2022, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- [9] I. U. W. Mulyono, E. H. Rachmawanto, C. A. Sari, and M. K. Sarker, "A high accuracy of deep learning based CNN architecture: classic, VGGNet, and ResNet50 for Covid-19 image classification," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 22, no. 5, pp. 1187–1195, Oct. 2024, doi: 10.12928/TELKOMNIKA.v22i5.26017.
- [10] W. Wijaya Kusuma, R. Rizal Isnanto, A. Fauzi, and P. Korespondensi, "DenseNet121 Menggunakan Kerangka Kerja TensorFlow untuk Deteksi Jenis Hewan," *Jurnal Teknik Komputer*, vol. 1, no. 4, pp. 141–147, 2023, doi: 10.14710/jtk.v1i4.37009.
- [11] F. Sultana, A. Sufian, and P. Dutta, "Advancements in Image Classification using Convolutional Neural Network," *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pp. 1–9, May 2019, doi: 10.1109/ICRCICN.2018.8718718.
- [12] M. A. Al-masni, D. H. Kim, and T. S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput Methods Programs Biomed*, vol. 190, pp. 1–2, Jul. 2020, doi: 10.1016/j.cmpb.2020.105351.
- [13] G. P. H. P. Gusti, E. Haerani, F. Syafria, F. Yanto, and S. K. Gusti, "Implementasi Algoritma Convolutional Neural Network (Resnet-50) untuk Klasifikasi Kanker Kulit Benign dan Malignant," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 984–992, Jun. 2024, doi: 10.57152/malcom.v4i3.1398.
- [14] S. Sa'idah, I. Putu, Y. Nugraha Suparta, and E. Suhartono, "Modifikasi Convolutional Neural Network Arsitektur GoogLeNet dengan Dull Razor Filtering untuk Klasifikasi Kanker Kulit," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 2, pp. 148–153, 2022.
- [15] D. Hindarto, "Comparison of Detection with Transfer Learning Architecture RestNet18, RestNet50, RestNet101 on Corn Leaf Disease," <https://jtiulm.ti.ft.ulm.ac.id/index.php/jtiulm/article/view/174>, vol. 8, no. 2, pp. 41–48, 2023, doi: <https://doi.org/10.20527/jtiulm.v8i2.174>.
- [16] M. Muzahidul Islam Rahi, A. K. M Amanat Ullah, F. Tanvir Khan, M. Golam Rabiul Alam, M. Tanvir Mahtab, and M. Ashrafal Alam, "DETECTION OF SKIN CANCER USING DEEP NEURAL NETWORKS," *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2019, doi: <https://doi.org/10.1109/CSDE48274.2019.9162400>.
- [17] D. A. Nurlitasari, R. Magdalena, and R. Y. N. Fu'adah, "ANALISIS PERFORMANSI SISTEM KLASIFIKASI KANKER KULIT MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK," *JOURNAL OF ELECTRICAL AND SYSTEM CONTROL ENGINEERING*, vol. 5, no. 2, pp. 91–99, Feb. 2022, doi: 10.31289/jesce.v5i2.5691.

SENASTIKA Universitas Malikussaleh

SEGMENTASI RESIKO OBESITAS DAN PENYAKIT JANTUNG DENGAN ALGORITMA K-MEANS

Muhammad Dzikrillah¹, Nuzul Althaf², Sylva Putri Utari³, Suhaiba Nasyira Hariono*⁴, Munirul Ula*⁵

1,2,3,4 Teknik Informatika, Universitas Malikussaleh, lhokseumawe. Aceh

Email: ¹muhammad.210170249@mhs.unimal.ac.id, ²nuzul.210170258@mhs.unimal.ac.id, ³sylva.210170247@mhs.unimal.ac.id, ⁴suhaiba.210170253@mhs.unimal.ac.id, ⁵munirulula@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan segmentasi risiko obesitas dan penyakit jantung dengan menerapkan algoritma K-Means dan mengidentifikasi faktor-faktor yang memengaruhi tingkat obesitas pada pasien pria dan wanita. Faktor-faktor yang dianalisis mencakup usia, tinggi badan, berat badan, riwayat keluarga dengan berat badan berlebih, frekuensi konsumsi makanan tinggi kalori, frekuensi konsumsi sayuran, jumlah makan utama dalam sehari, kebiasaan ngemil, kebiasaan merokok, asupan air, pemantauan kalori, frekuensi aktivitas fisik, durasi penggunaan perangkat teknologi, konsumsi alkohol, dan jenis transportasi yang digunakan. Hasil penelitian menunjukkan bahwa analisis terbaik dari setiap metode dan optimasi K-Means menghasilkan nilai Davies-Bouldin Index optimal pada $n_clusters = 3$ dengan nilai DBI 1.6545, serta nilai terbaik untuk Silhouette Coefficient pada $n_clusters = 4$ dengan rata-rata 0.2445. Temuan ini menunjukkan bahwa K-Means merupakan metode yang paling efektif untuk segmentasi risiko obesitas dan penyakit jantung. Meskipun nilai evaluasi clustering mungkin tidak optimal, hal ini tidak selalu mencerminkan kualitas pemodelan; bisa jadi ini menunjukkan bahwa data tidak terkelompok dengan jelas atau perlu penyesuaian parameter model. Oleh karena itu, evaluasi sebaiknya digabungkan dengan pemahaman tentang domain dan analisis eksploratif untuk memastikan model clustering sesuai dengan kebutuhan analisis.

Keywords: *Segmentasi, Risiko Obesitas, Penyakit Jantung, Algoritma K-Means, Silhouette Coefficient dan Davies-Bouldin Index*

1. PENDAHULUAN

Obesitas telah menjadi masalah kesehatan yang signifikan dengan prevalensi yang meningkat. Diperkirakan 1,6 miliar orang dewasa di dunia mengalami kelebihan berat badan, dan 400 juta di antaranya obesitas. Pada tahun 2015, diperkirakan 2,3 miliar orang akan overweight dan 700 juta diantaranya obesitas. Penyakit jantung terjadi saat aliran darah ke otot jantung tersumbat, menyebabkan jantung tidak berfungsi dengan baik dan berpotensi mengalami kerusakan serius atau penyempitan arteri koroner[1]. Faktor risiko terbagi menjadi yang tidak dapat diubah, seperti usia, jenis kelamin, dan keturunan, serta yang dapat diubah, seperti hipertensi, kolesterol tinggi, obesitas, diabetes, kurang aktivitas fisik, dan konsumsi alkohol berlebih. Menurut peneliti Haryadi dkk, [2] Serangan jantung merupakan penyakit mematikan dengan angka kematian mencapai 12,90% dari semua penyakit jantung.

Kurangnya akses informasi tentang penyakit jantung menyebabkan peningkatan angka kematian setiap tahun, sehingga dibutuhkan sistem klasifikasi untuk memberikan informasi dan melakukan pengecekan dini[3]. Obesitas umumnya disebabkan oleh ketidakseimbangan antara asupan makanan dan aktivitas fisik, dengan individu obesitas cenderung makan berlebihan dan kurang berolahraga. Namun, obesitas dipengaruhi oleh berbagai faktor, termasuk genetik, fisiologis, metabolik, psikologis, sosioekonomik, gaya hidup, dan budaya[4]. Obesitas berisiko meningkatkan penyakit kronis seperti hipertensi, diabetes, dan kanker, serta dapat menimbulkan stigma sosial bagi penderitanya.

Penyakit jantung dan obesitas menjadi dua isu kesehatan yang saling terkait dan semakin mendesak untuk diatasi dalam masyarakat modern[5]. Dengan meningkatnya angka kematian akibat penyakit jantung, terutama yang disebabkan oleh kurangnya akses informasi dan pemahaman yang tepat mengenai kondisi ini, muncul kebutuhan mendesak akan sistem klasifikasi yang dapat membantu dalam deteksi dini dan penanganan penyakit. Di sisi lain, obesitas yang sering kali berakar dari pola hidup yang tidak seimbang, menjadi pemicu berbagai penyakit kronis yang berkontribusi pada meningkatnya risiko kematian[6].

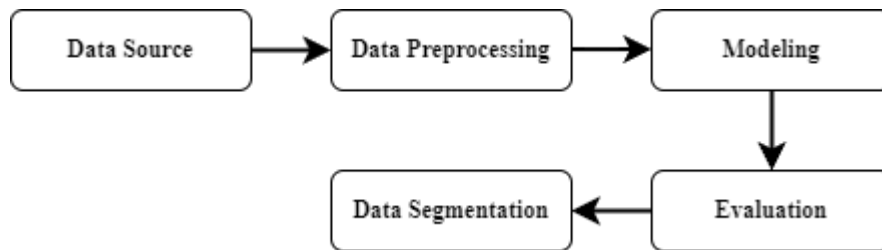
Penyakit jantung dan obesitas menjadi dua isu kesehatan yang saling terkait dan semakin mendesak untuk diatasi dalam masyarakat modern. Dengan meningkatnya angka kematian akibat penyakit jantung, terutama yang disebabkan oleh kurangnya akses informasi dan pemahaman yang tepat mengenai kondisi ini, muncul kebutuhan mendesak akan sistem klasifikasi yang dapat membantu dalam deteksi dini dan penanganan penyakit. Di sisi lain, obesitas yang sering kali berakar dari pola hidup yang tidak seimbang menjadi pemicu berbagai penyakit kronis yang berkontribusi pada meningkatnya risiko kematian. Pemahaman yang lebih mendalam tentang pola obesitas

dan faktor-faktor risikonya diperlukan untuk mendukung upaya pencegahan dan pengobatan yang lebih efisien. Salah satu pendekatan yang bisa digunakan dalam analisis data kesehatan adalah metode pengelompokan (clustering)[7].

Tujuan dari penelitian ini adalah untuk mengidentifikasi faktor-faktor penyebab obesitas dan penyakit jantung, serta mengeksplorasi hubungan antara keduanya[8]. Penelitian ini menggunakan algoritma K-Means dengan evaluasi hasil clustering melalui Silhouette Score dan Davies-Bouldin Index[9]. Kedua metode evaluasi internal ini membantu mengukur kualitas pengelompokan data berdasarkan karakteristik risiko dan dampak sosial yang dihadapi individu dengan kondisi tersebut. Silhouette Score mengukur seberapa baik data dalam sebuah cluster dibandingkan dengan cluster lainnya, sementara Davies-Bouldin Index mengevaluasi jarak antar cluster, memberikan gambaran tentang homogenitas dan kualitas cluster yang dihasilkan[10]. Beberapa penelitian sebelumnya mendukung penggunaan metrik ini dalam mengevaluasi hasil clustering. Penelitian ini diharapkan memberikan kontribusi signifikan dengan mengusulkan penggunaan Silhouette Score dan Davies-Bouldin Index sebagai metrik evaluasi untuk menilai efektivitas clustering K-Means, yang dapat bermanfaat bagi masyarakat dalam upaya pencegahan obesitas dan penyakit jantung.

2. METODE PENELITIAN

Tahapan penelitian merupakan rangkaian langkah terstruktur yang diterapkan agar proses penelitian menjadi lebih mudah dan hasilnya tersusun dengan baik. Pada gambar 1 menggambarkan tahapan-tahapan penelitian.



Gambar 1. Tahapan Penelitian

Gambar 1 memperlihatkan tahapan penelitian, dimulai dari pengambilan dataset pada tahap Data Source untuk memulai penelitian. Selanjutnya, pada tahap Data Preprocessing, dataset diolah agar siap untuk dianalisis dan dimodelkan. Pada tahap ketiga, dilakukan Modeling dengan menerapkan algoritma K-Means pada dataset yang telah diproses. Kemudian, pada tahap Evaluation, evaluasi dilakukan terhadap cluster yang telah terbentuk. Terakhir, pada tahap Data Segmentation, cluster terbaik dipilih dari semua cluster yang telah dihasilkan.

a. Data Source

Data yang digunakan dalam penelitian ini berasal dari dataset yang dapat diakses di situs web data publik dengan mencari menggunakan kata kunci “Obesity/CVD Risk Clustering with K-Means” di platform Kaggle.com.

Tabel 1. Dataset Penyakit Obesitas Dan Jantung

No	Gender	Age	V1	V2	...	V11	V12	V13	V14	V15	V16	V17
1	female	21	1,62	64	...	2	no	0	1	no	public_trnsprt	normal_weight
2	female	21	1,52	56	...	3	yes	3	0	sometimes	public_trnsprt	normal_weight
3	male	23	1,8	77	...	2	no	2	1	frequently	public_trnsprt	normal_weight
4	male	27	1,8	87	...	2	no	2	0	frequently	walking	overweight_level_i
...
2106	female	21,28	1,72	131,3		1,79	no	1,72	0,89	sometimes	public_trnsprt	obesity_type_iii
2107	female	20,97	1,71	131,4		1,72	no	1,67	0,90	sometimes	public_trnsprt	obesity_type_iii
2108	female	21,98	1,74	133,7		2,00	no	1,34	0,59	sometimes	public_trnsprt	obesity_type_iii
2119	female	22,52	1,75	133,6		2,05	no	1,41	0,64	sometimes	public_trnsprt	obesity_type_iii
2110	female	24,36	1,73	133,3		2,85	no	1,13	0,58	sometimes	public_trnsprt	obesity_type_iii
2111	female	23,66	1,73	133,4		2,86	no	1,02	0,71	sometimes	public_trnsprt	obesity_type_iii

b. Data Preprocessing

Data Preprocessing adalah proses pengolahan data agar dapat dipahami oleh mesin dengan lebih mudah [11]. Proses ini mencakup beberapa langkah, seperti:

1. Memilih atribut yang relevan untuk digunakan dalam modeling dengan algoritma k-means.
2. Mengisi nilai yang hilang dalam dataset menggunakan rata-rata.
3. Menghapus data yang duplikat dari dataset.
4. Memeriksa atribut yang mengindikasikan adanya outlier dan melakukan normalisasi data [12].

c. Modeling

Dengan menentukan nilai parameter yang optimal, tahap modeling ini mengimplementasikan algoritma k-means. K-Means adalah algoritma klasterisasi yang membagi objek pengamatan ke dalam kelompok menggunakan metode partisi [13]. Algoritma ini membagi data ke dalam kelompok berdasarkan titik pusat (centroid). Langkah-langkah dalam perhitungan algoritma k-means adalah sebagai berikut:

1. Tentukan jumlah cluster (k) yang diinginkan.
2. Pilih centroid cluster secara acak.
3. Hitung jarak setiap data terhadap centroid menggunakan rumus Euclidean:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1)$$

Di mana $d(x, y)$ melambangkan jarak antara dua titik data; x adalah titik data pada objek, dan y adalah titik pusat. Jumlah objek dilambangkan dengan i .

4. Kelompokkan data berdasarkan jarak minimum antara centroid dan data untuk membentuk cluster baru.
5. Hitung ulang centroid (pusat cluster baru) untuk setiap cluster menggunakan rumus:

$$CI = \frac{x1 + x2 + x3 + \dots + xn}{\sum x} \quad (2)$$

CI adalah titik pusat baru dari sebuah cluster, dengan $x1$ hingga xn mewakili nilai dalam cluster ke- n .

6. Ulangi langkah 3 hingga 5 hingga tidak ada data yang berpindah cluster.

Sebelum mengimplementasikan algoritma k-means [14], perlu dilakukan penentuan jumlah cluster yang sesuai. Metode Elbow digunakan untuk mengukur konsistensi jumlah cluster yang optimal, yang ditandai dengan adanya lekukan pada grafik yang menyerupai siku.

d. Evaluation

Evaluation adalah proses pengukuran hasil dari model yang telah diimplementasikan pada tahap modeling sebelumnya. Pada tahap ini, dilakukan penilaian terhadap tingkat keberhasilan dan akurasi cluster yang dihasilkan dari proses modeling. Beberapa metode evaluasi yang digunakan antara lain Silhouette Score, dan Davies Bouldin Index (DBI).

e. Data Segmentation

Data segmentation digunakan untuk mengelompokkan individu berdasarkan faktor-faktor yang memengaruhi risiko obesitas dan penyakit jantung, seperti usia, berat badan, riwayat keluarga, pola makan, kebiasaan merokok, dan aktivitas fisik. Dengan algoritma K-Means, individu dikelompokkan ke dalam segmen dengan karakteristik serupa, sehingga dapat memudahkan identifikasi kelompok berisiko tinggi untuk intervensi yang lebih tepat dalam pencegahan dan pengobatan.

3. HASIL DAN PEMBAHASAN

3.1 Data Source

Data Source dalam penelitian ini menggunakan dataset Obesity/CVD Risk. Tabel 2 menjelaskan setiap atribut yang terdapat dalam dataset tersebut.

Tabel 2. Deskripsi *Dataset Obesity-CVD Risk*

Atribut	Keterangan	Penjelasan
Gender	Jenis kelamin	Male= 51% dan female = 49%
Age	Umur(tahun)	Umur pasien “min = 14, dan maks = 61”.
Height	Tinggi Badan	Tinggi badan pasien “min = 1,45 cm dan maks = 1,98 cm”.
Weight	Berat Badan	Berat badan pasien “min = 39 kg, dan maks = 173 kg”.
FHWO	family_history_with_overweight	Memiliki Keluarga yang berat badannya tinggi dengan kategori “yes = 82% dan no 18%”.
FAVC	Frequent consumption of high caloric food	Frekuensi mengonsumsi makanan berkalori tinggi dengan kategori “yes = 88% dan no = 12%”.
FCVC	Frequency of consumption of vegetables	Frekuensi Konsumsi Sayur perhari dengan kategori 1(sedikit), 2(sedang) dan 3(banyak).
NCP	Number of main meals	Jumlah makan dalam sehari dengan kategori 1(sedikit), 2(sedang) dan 3(banyak).
CAEC	Consumption of food between meals	Konsumsi makanan atau cemilan di antara waktu makan dengan kategori “no = 5%, sometimes = 84%, frequently = 11% dan always = 5%”
SMOKE	Merokok	pasien dibedakan dengan kategori “no dan yes”
CH20	Consumption of water daily	Jumlah Air yang diminum dalam sehari dengan kategori 1(sedikit), 2(sedang) dan 3(banyak).
SCC	Calories consumption monitoring	Pemantauan konsumsi jumlah kalori dengan kategori “yes dan no”
FAF	Physical activity frequency	Frekuensi aktivitas fisik dengan kategori 0(tidak), 1(jarang), 2(sering) dan 3(selalu).
TUE	Time using technology devices	Menghabiskan waktu menggunakan perangkat teknologi dengan kategori 0(tidak), 1(jarang) dan 2(selalu).
CALC	Consumption of alcohol	Jumlah alkohol yang dikonsumsi dengan kategori no(tidak), sometimes(kadang), frequently(sering), dan always(selalu)
MTRANS	Transportation used	Transportasi yang digunakan sehari hari pasien dengan kategori “Public transportation, automobile, motorbike, bike, dan walking”.
Nobeyesdad	Tingkat obesitas dikurangi	Mengacu pada kategori atau tingkat obesitas yang ditentukan seperti: Insufficient_Weight, Normal_Weight, Overweight_Level_I, Overweight_Level_II, Obesity_Type_I, Obesity_Type_II, Obesity_Type_III

Dataset ini mencakup data untuk estimasi tingkat obesitas pada individu berdasarkan kebiasaan makan dan kondisi fisik mereka. Data berisi 17 atribut dan 500 catatan, catatan diberi label dengan variabel kelas NObesity (Tingkat Obesitas), yang memungkinkan klasifikasi / pengelompokan / Regresi data menggunakan nilai Berat Badan Kurang, Berat Badan Normal, Berat Badan Berlebih I, Berat Badan Berlebih II, Jenis Obesitas I, Jenis Obesitas II, dan Jenis Obesitas III. Setelah menggunakan algoritma K-Means untuk klusterisasi dataset, kami menemukan bahwa tiga klaster memberikan segmentasi terbaik dalam menggambarkan risiko obesitas individu. Klaster-klaster tersebut dibagi berdasarkan faktor gaya hidup seperti frekuensi aktivitas fisik, kebiasaan mengonsumsi makanan berkalori tinggi, serta faktor genetik seperti riwayat keluarga dengan obesitas, dll. Setelah

mengetahui penjelasan atribut pada dataset kemudian melanjutkan pada tahap *data processing* untuk dianalisis lebih lanjut.

3.2 Data Preprocessing

Sebelum penerapan algoritma K-Means, dilakukan praproses data untuk memastikan kualitas data yang baik [11]. tahapan praproses meliputi:

1. Konversi data kategorikal menjadi numerik: Karena K-means menggunakan jarak Euclidean, data kategorikal perlu diubah ke format numerik[15]. Ini dapat dilakukan dengan metode seperti Label Encoding atau One-Hot Encoding. Data ini terdapat pada tabel 3 dibawah ini.

Tabel 3. Konfersi Data Kategorikal-Numerik

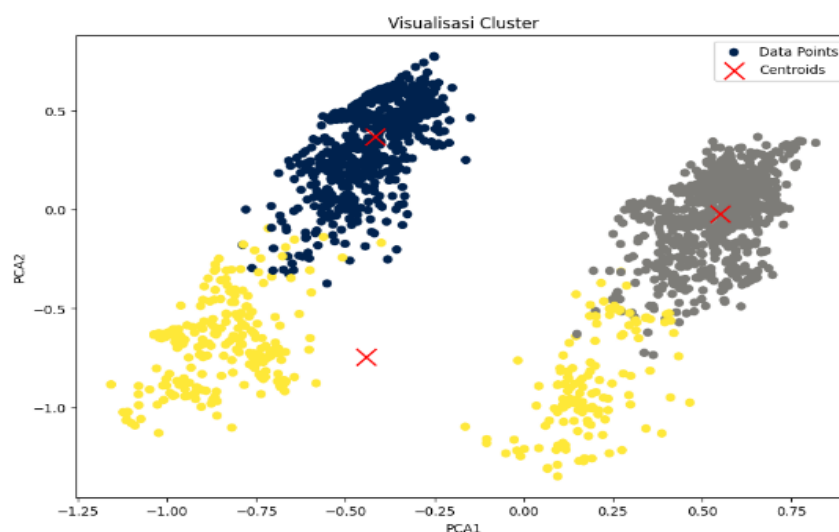
Index	Gender	Age	V1	V2	...	V13	V14	V15	V16	V17
0	0	21	1,62	64	...	0	1	3	3	1
1	0	21	1,52	56	...	3	0	2	3	1
2	1	23	1,8	77	...	2	1	1	3	1
3	1	27	1,8	87	...	2	0	1	4	5
4	1	22	1,78	89,8	...	0	0	2	3	6

2. Normalisasi data: Data numerik sering kali perlu dinormalisasi untuk memastikan skala yang konsisten. Hal ini penting karena K-means peka terhadap skala variabel[12]. Salah satu metode untuk melakukan ini adalah MinMaxScaler, yang mengubah data sehingga fitur-fitur berada dalam rentang [0, 1]. Normalisasi data merupakan proses penyesuaian skala data agar semua fitur memberikan kontribusi yang seimbang terhadap algoritma k-means dan berikut ini hasil setelah di normalisasi.

Tabel 4. Hasil Normalisasi Data

Index	Gender	Age	V1	V2	...	V13	V14	V15	V16	V17
0	0	0,14	0,32	0,18	...	0	0,5	1	0,75	0,16
1	0	0,14	0,13	0,12	...	1	0	0,66	0,75	0,16
2	1	0,19	0,66	0,28	...	0,66	0,5	0,33	0,75	0,16
3	1	0,27	0,66	0,35	...	0,66	0	0,33	1	0,83
4	1	0,17	0,62	0,37	...	0	0	0,66	0,75	1

3. Visualisasi cluster: Setelah model di-fit, hasil klusterisasi dapat divisualisasikan menggunakan plot dua dimensi jika data telah direduksi, atau dengan teknik visualisasi multidimensi lainnya.



Gambar 2. Hasil visualisasi cluster

Centroid dari setiap kluster, yang merupakan titik tengah dari kluster, dapat ditampilkan pada visualisasi untuk memberikan pemahaman tentang pusat setiap kluster.

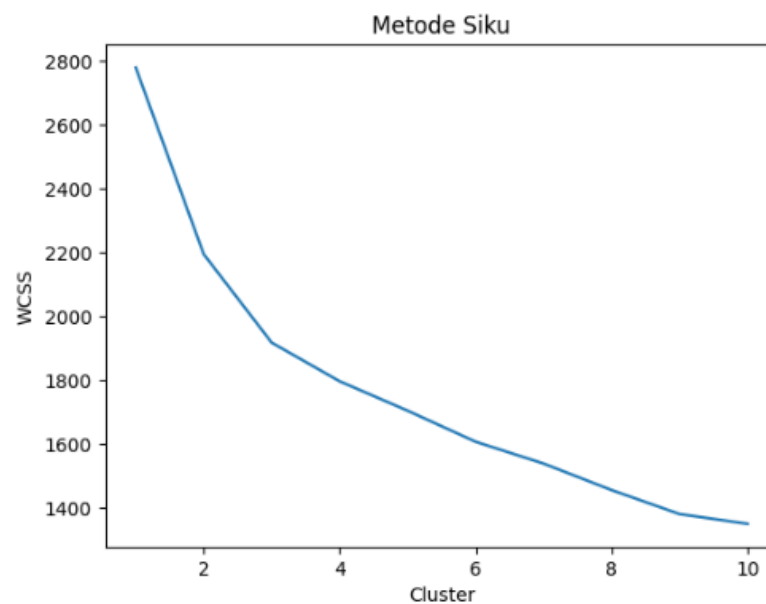
3.3 Modelling

Setelah menyelesaikan semua langkah dalam data preprocessing, dataset siap untuk menentukan model tingkat risiko terkena obesitas dan penyakit jantung yang dikategorikan sebagai "0 = Rendah, 1 = Sedang, dan 2 = Tinggi". pada tabel dibawah ini

Tabel 5. Hasil Klasterisasi

Gender	Age	V1	V2	...	V13	V14	V15	V16	V17	Cluster	Resiko Terkena(Cluster Label)
Female	21	1,62	64	...	0	1	no	Public_Transport	Nrml_Weight	0	Rendah
Female	21	1,52	56	...	3	0	Sometimes	Public_Transport	Nrml_Weight	0	Rendah
Male	23	1,8	77	...	2	1	Frequently	Public_Transport	Nrml_Weight	1	Sedang
Male	27	1,8	87	...	2	0	Frequently	Walking	Overweight_Lvl_I	2	Tinggi
...
Female	21,28	1,72	131,33	...	1,72	0,89	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	20,97	1,71	131,40	...	1,67	0,90	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	21,98	1,74	133,74	...	1,34	0,59	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	22,52	1,75	133,68	...	1,41	0,64	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	24,36	1,73	133,34	...	1,13	0,58	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	23,66	1,73	133,47	...	1,02	0,71	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah

Kemudian untuk menentukan jumlah kluster yang optimal, metode elbow digunakan. Ini melibatkan plot dari jumlah kluster terhadap inersia (total jarak dalam kluster) dan mencari titik "siku" di mana peningkatan jumlah kluster tidak lagi memberikan penurunan inersia yang signifikan.

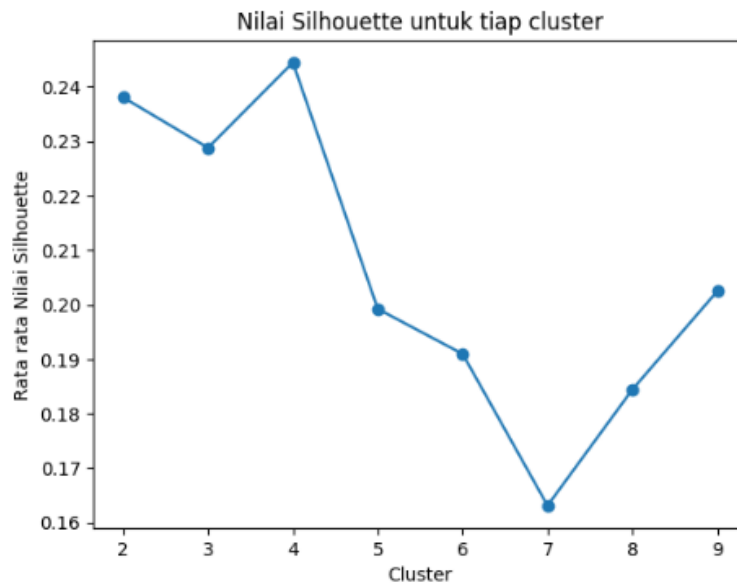


Gambar 3. Grafik Elbow Method

Gambar 3 memperlihatkan grafik elbow method yang menunjukkan penurunan signifikan pada WCSS hingga mencapai sekitar 3 cluster, setelah itu penurunan menjadi lebih lambat. Oleh karena itu, maka nilai $k = 3$ ditetapkan sebagai nilai optimal untuk pengolahan data menggunakan algoritma K-Means.

3.4 Evaluation

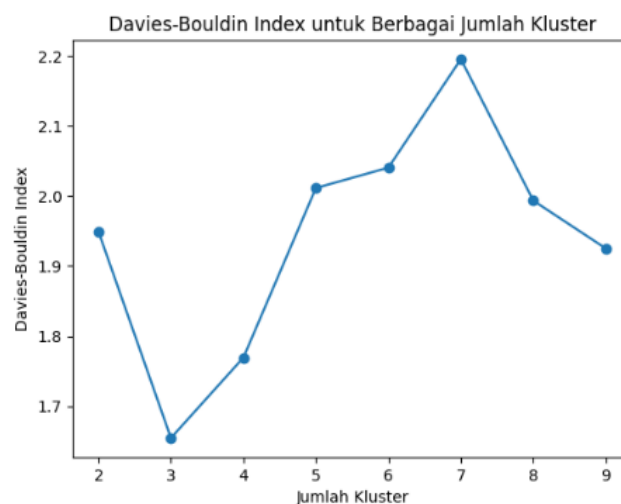
a. Silhouette Score



Gambar 4. Grafik Nilai Silhouette Score

Silhouette score mengukur kemiripan setiap titik data dengan kelompoknya dibandingkan dengan kelompok terdekat[16]. Nilai yang lebih tinggi menunjukkan pengelompokan yang lebih baik, di mana titik data lebih dekat ke centroid kelompoknya dan lebih jauh dari centroid kelompok lain[17]. Gambar 6 menunjukkan bahwa cluster 2 dan 6 memiliki nilai silhouette score yang cukup tinggi.

b. Davies Bouldin Index (DBI)



Gambar 5. Grafik Nilai Davies Bouldin Index

Davies Bouldin Index mengukur rasio antara jarak rata-rata intra-cluster (jarak titik data dalam kelompok) dan jarak minimum antar centroid kelompok. Nilai yang lebih rendah menunjukkan pengelompokan yang lebih baik[18]. Gambar 7 menunjukkan bahwa cluster ke-3 memiliki nilai Davies Bouldin Index terendah.

3.5 Data Segmentation

Pada tahap data segmentation, yaitu menentukan cluster mana berdasarkan dari dataset diatas, peneliti mengidentifikasi dan mendapatkan sejumlah faktor yang memengaruhi tingkat obesitas pada pasien pria dan wanita dengan penyakit jantung. Faktor-faktor tersebut mencakup usia, tinggi badan, berat badan, riwayat keluarga dengan berat badan berlebih (*family_history_with_overweight*), frekuensi konsumsi makanan tinggi kalori (FAVC), frekuensi konsumsi sayuran (FCVC), jumlah makan utama dalam sehari (NCP), kebiasaan ngemil (CAEC), kebiasaan merokok (SMOKE), jumlah air yang diminum setiap hari (CH2O), pemantauan konsumsi kalori (SCC), frekuensi aktivitas fisik (FAF), durasi penggunaan perangkat teknologi (TUE), konsumsi alkohol (CALC), transportasi yang digunakan sehari-hari (MTRANS), serta tingkat obesitas yang dikurangi (NObesidad).

3. DISKUSI

Hasil penelitian kami menunjukkan bahwa algoritma K-Means efektif dalam melakukan segmentasi risiko obesitas dan penyakit jantung dengan mempertimbangkan berbagai faktor seperti pola konsumsi makanan, aktivitas fisik, dan kebiasaan hidup sehari-hari. Temuan ini sejalan dengan jurnal terdahulu yang juga menggunakan algoritma K-Means untuk mengelompokkan makanan cepat saji berdasarkan kandungan nutrisinya guna membantu penderita obesitas memilih makanan yang lebih sehat.

Namun, terdapat beberapa perbedaan dalam pendekatan dan hasil penelitian. Jurnal terdahulu yang berjudul **"Implementasi K-Means Untuk Pengelompokan Makanan Cepat Saji Bagi Penderita Penyakit Obesitas"** yang ditulis oleh Yoannes Dion Pradvenanta dan Rastri Prathivi yang dipublikasikan pada bulan Juni Tahun 2024 berfokus pada segmentasi makanan cepat saji bagi penderita obesitas, di mana data diambil dari 515 sampel makanan, dan hasil segmentasi menunjukkan bahwa dari 5 cluster yang dihasilkan, 244 sampel makanan dapat dianggap sebagai pilihan yang lebih sehat. Di sisi lain, penelitian kami lebih menitikberatkan pada segmentasi risiko obesitas dan penyakit jantung pada individu dengan dataset yang mencakup 2111 sampel pasien. Dengan menggunakan 3 hingga 4 cluster, hasil penelitian kami memberikan pemahaman yang lebih komprehensif mengenai faktor risiko yang terkait dengan obesitas dan penyakit jantung, termasuk riwayat keluarga, pola makan tinggi kalori, dan frekuensi aktivitas fisik.

Dalam hal prevalensi obesitas, perbandingan data dari kedua jurnal menunjukkan tren yang mengkhawatirkan. Berdasarkan jurnal kami, terdapat 1,6 miliar orang dewasa di dunia yang mengalami kelebihan berat badan, dengan 400 juta di antaranya menderita obesitas parah. Angka ini diperkirakan terus meningkat menjadi 2,3 miliar orang yang mengalami kelebihan berat badan dan 700 juta dengan obesitas pada tahun 2015. Sementara itu, jurnal terdahulu melaporkan bahwa pada tahun 2022, satu dari delapan orang di dunia menderita obesitas, termasuk 890 juta orang dewasa dan 160 juta anak-anak dan remaja. Hal ini mempertegas bahwa obesitas menjadi salah satu masalah kesehatan global yang semakin serius, dengan dampak signifikan terhadap risiko penyakit tidak menular, termasuk penyakit jantung, diabetes, dan gangguan metabolik lainnya.

Evaluasi model yang digunakan pada kedua penelitian juga memperlihatkan hasil yang berbeda. Penelitian kami menunjukkan bahwa nilai terbaik untuk Davies-Bouldin Index (DBI) adalah 1.6545 pada 3 cluster, dan Silhouette Coefficient sebesar 0.2445 pada 4 cluster, yang menunjukkan pemisahan cluster yang cukup baik. Di sisi lain, jurnal terdahulu menggunakan Sum Square Error (SSE), Silhouette Score, dan DBI untuk menilai clustering makanan cepat saji, dengan hasil optimal pada 5 cluster.

Perbandingan ini menunjukkan bahwa meskipun kedua studi menggunakan algoritma yang sama, fokus analisis dan populasi yang diteliti berbeda. Penelitian kami lebih berfokus pada identifikasi faktor risiko dan segmentasi pasien, sedangkan jurnal terdahulu lebih fokus pada segmentasi makanan berdasarkan kandungan nutrisinya untuk membantu penderita obesitas membuat pilihan yang lebih sehat. Kedua pendekatan ini saling melengkapi dalam upaya mengurangi risiko obesitas dan penyakit terkait.

4. KESIMPULAN

Dari penelitian ini dapat disimpulkan bahwa algoritma K-Means efektif untuk segmentasi risiko obesitas dan penyakit jantung berdasarkan berbagai faktor. Tiga cluster teridentifikasi: Cluster 0 (risiko rendah), Cluster 1 (risiko sedang), dan Cluster 2 (risiko tinggi). Usia juga berpengaruh, di mana individu yang lebih tua cenderung memiliki risiko lebih tinggi. Temuan ini menunjukkan pentingnya pemantauan faktor risiko dan perlunya pendekatan terfokus dalam pencegahan. Nilai evaluasi yang tidak optimal tidak selalu mencerminkan buruknya pemodelan, tetapi bisa jadi karena kurangnya pengelompokan jelas dalam data atau perlunya penyesuaian parameter. Oleh karena itu, evaluasi harus dikombinasikan dengan pemahaman domain dan analisis eksploratif untuk memastikan kesesuaian model clustering dengan kebutuhan analisis.

DAFTAR PUSTAKA

- [1] F. Rahmadayanti, S. Muntari, dan R. Putriani, “PENGELOMPOKAN TINGKAT RESIKO PENYAKIT JANTUNG BERDASARKAN USIA MENGGUNAKAN ALGORITMA K-MEANS,” 2023.
- [2] D. Haryadi dan D. Marini Umi Atmaja, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung,” *Journal of Informatics and Communications Technology*, vol. 3, no. 2, hlm. 051–066, 2021.
- [3] J. Wala dan R. Umar, “Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung,” 2024.
- [4] H. H. Satoto*, “TINJAUAN PUSTAKA Patofisiologi Penyakit Jantung Koroner Coronary Heart Disease Pathophysiology,” 2014.
- [5] Sonia Indhira dan Billy Hendrik, “Implementation Of The K-Means Clustering Algorithm For Grouping Heart Disease Risk Levels,” *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 3, no. 3, hlm. 154–170, Okt 2023, doi: 10.55606/juisik.v3i3.677.
- [6] Y. Dion Pradvenanta dan R. Prathivi, “Implementasi K-Means Untuk Pengelompokan Makanan Cepat Saji Bagi Penderita Penyakit Obesitas,” *Technology and Science (BITS)*, vol. 6, no. 1, 2024, doi: 10.47065/bits.v6i1.5279.
- [7] R. Wahyusari, S. Tinggi, dan T. Ronggolawe, “Penerapan Algoritma K-Medoids Untuk Mengelompokkan Status Obesitas,” 2024.
- [8] I. A. T. H. A. B. S. Ice J. Johanis1, “Media Kesehatan Masyarakat FAKTOR RISIKO HIPERTENSI, MEROKOK DAN USIA TERHADAP,” *Media Kesehatan Masyarakat*, vol. 2, no. 1, 2020, [Daring]. Tersedia pada: <https://ejurnal.undana.ac.id/MKM>
- [9] F. A. B. N. S. Abyan Rasyid Aryadi, “Analisis Klaster K-Means pada Data Rata-Rata Konsumsi Kalori dan Protein Menurut Provinsi dengan Metode Davies Bouldin Index,” 2022.
- [10] D. Hartama dan I. Sudahri Damanik, “Pengelompokan Algoritma K-Means dan K-Medoid Berdasarkan Lokasi Daerah Rawan Bencana di Indonesia dengan Optimasi Elbow, DBI, dan Silhouette,” *Technology and Science (BITS)*, vol. 6, no. 2, 2024, doi: 10.47065/bits.v6i2.5851.
- [11] G. Fouriza Ibanez, W. Giri, dan R. Wiriasto, “Kombinasi Principal Component Analysis dengan Algoritma K-Means untuk Klasterisasi Data Stunting,” *Media Online*, vol. 5, no. 1, hlm. 131–141, 2024, doi: 10.30865/klik.v5i1.1977.
- [12] L. Mutawalli, S. Fadli, dan S. Supardianto, “Komparasi Metode Perhitungan Jarak K-Means Paling Baik Terhadap Pembentukan Pola Kunjungan Wisatawan Mancanegara,” *Journal of Information System Research (JOSH)*, vol. 5, no. 1, hlm. 159–166, Okt 2023, doi: 10.47065/josh.v5i1.4377.
- [13] Y. R. Nasution dan M. Eka, “PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA APLIKASI MENENTUKAN BERAT BADAN IDEAL,” 2018.
- [14] A. Satriawan, R. Andreswari, dan O. N. Pratiwi, “SEGMENTASI PELANGGAN TELKOMSEL MENGGUNAKAN METODE CLUSTERING DENGAN RFM MODEL DAN ALGORITMA K-MEANS TELKOMSEL CUSTOMER SEGMENTATION USING CLUSTERING METHOD WITH RFM MODEL AND K-MEANS ALGORITHM,” 2021.
- [15] A. Saputra dan R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, hlm. 1346–1361, Jul 2024, doi: 10.57152/malcom.v4i4.1516.
- [16] K. Amrulloh, T. Hendro Pudjiantoro, P. Nurul Sabrina, dan A. Id Hadiana, “Comparison Between Davies-Bouldin Index and Silhouette Coefficient Evaluation Methods in Retail Store Sales Transaction Data Clusterization Using K-Medoids Algorithm,” 2022. [Daring]. Tersedia pada: www.kaggle.com,
- [17] Y. Hasan, “Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan,” 2024.

- [18] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, dan R. Liwardana, “Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta,” 2023. [Daring]. Tersedia pada: <http://jurnal.polibatam.ac.id/index.php/JAIC>

SENASTIKA Universitas Malikussaleh

DETEKSI DINI PENYAKIT KATARAK PADA MATA MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN)

Muhammad Khardawi^{*1}, Aridho Pangestu², Fajar Satria³, Muhammad Naufal Hadi⁴, Munirul Ula⁵

^{1,2,3,4,5} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: ¹muhammad.210170255@mhs.unimal.ac.id, ²aridho.210170266@mhs.unimal.ac.id,
³fajar.210170260@mhs.unimal.ac.id, ⁴muhammad.210170264@mhs.unimal.ac.id, ⁵munirulula@unimal.ac.id

Abstrak

Mata adalah organ vital dalam manusia yang berfungsi sebagai organ penglihatan. Satu hal yang sangat penting untuk dipertimbangkan adalah masalah kebutaan. Katarak memiliki dampak signifikan terhadap kualitas hidup penderita dan dapat berkontribusi pada peningkatan kemiskinan di Indonesia. Menurut data dari World Health Organization Dunia (WHO), katarak menyumbang sekitar 48% kasus kebutaan global dengan prevalensi tinggi di Indonesia. Karena mata normal sulit dikenali dari mata penderita katarak, penderita sering kali tidak mengenali adanya katarak. Oleh karena itu, sangat penting untuk mengidentifikasi katarak yang baru saja terjadi gangguan penglihatan. Dengan kemajuan teknologi, bukti dan klasifikasi katarak yang dapat dikenali menjadi lebih mudah berkat persiapan gambar yang canggih. Dalam penelitian ini, penulis menggunakan metode convolutional neural network (CNN) untuk membuat aplikasi yang menginstruksikan machine learning dalam membedakan mata normal dan mata katarak. Dataset yang digunakan merupakan gambar dari kaggle dengan dua citra digital yaitu mata normal dan mata katarak. Pekerjaan pembelajaran mesin epoch 1, epoch 10, dan epoch 30. epoch 1 menghasilkan hasil dengan ketepatan akurasi sebesar 52%, epoch 10 menghasilkan hasil dengan akurasi 84% dan epoch 30 menghasilkan presisi tertinggi sebesar 87%. Munculnya pengujian visual menggunakan strategi CNN bisa sangat efektif dalam memprediksi mata katarak.

Kata kunci : *Citra Digital, Convolutional Neural Network (CNN), Katarak, Klasifikasi, Machine Learning.*

Abstract

The eye is a vital organ in humans, serving as the organ of vision. One crucial issue to consider is blindness. Cataracts have a significant impact on the quality of life for sufferers and can contribute to increasing poverty in Indonesia. According to data from the World Health Organization (WHO), cataracts account for around 48% of global blindness cases, with a high prevalence in Indonesia. Since normal eyes are difficult to distinguish from cataract-affected eyes, patients often do not realize they have cataracts. Therefore, it is essential to identify cataracts as soon as vision impairment begins. With advances in technology, evidence and classification of cataracts have become easier to recognize thanks to sophisticated image preparation. In this research, the author uses the Convolutional Neural Network (CNN) method to develop an application that instructs machine learning to differentiate between normal eyes and cataract-affected eyes. The dataset used consists of images from Kaggle, featuring two digital image categories: normal eyes and cataract-affected eyes. The machine learning process was performed at epoch 1, epoch 10, and epoch 30. Epoch 1 resulted in an accuracy of 52%, epoch 10 achieved 84%, and epoch 30 produced the highest precision at 87%. Visual testing using the CNN method proves to be highly effective in predicting cataracts.

Keywords : *Digital Image, Convolutional Neural Network (CNN), Cataract, Classification, Machine Learning.*

1. PENDAHULUAN

Mata merupakan organ vital pada manusia yang berperan dalam fungsi penglihatan. Salah satu aspek penting yang perlu diperhatikan adalah risiko kebutaan. Katarak menjadi penyebab utama kebutaan di Indonesia, yang berdampak signifikan terhadap kualitas hidup masyarakat. Berdasarkan data dari Organisasi Kesehatan Dunia (WHO), katarak menjadi faktor dominan penyebab kebutaan global, dengan sekitar 48% kasus terjadi di Indonesia. Kondisi ini turut berkontribusi terhadap kemiskinan di negara tersebut. [1]. Katarak atau kekeruhan pada lensa mata adalah salah satu penyebab kebutaan yang paling umum di Indonesia dan dunia. Diperkirakan, insiden katarak mencapai 0,1% per tahun, atau sekitar 1 dari 1.000 orang baru mengalami katarak setiap tahunnya. Penduduk Indonesia cenderung menderita katarak sekitar lima belas tahun lebih awal dibandingkan dengan penduduk di daerah subtropis. Sekitar 16-22% dari pasien katarak di Indonesia menjalani operasi sebelum mencapai usia lima puluh lima tahun. [2]. Katarak dapat menyebabkan denaturasi protein atau kombinasi keduanya, sekitar 90% kasus dikaitkan dengan usia, alasan lain adalah bawaan dan traumatis [3]. Seiring

bertambahnya usia, lensa mata mengalami kerusakan akibat paparan terus-menerus terhadap faktor eksternal, yang menyebabkan peningkatan radikal bebas. Radikal bebas ini bereaksi dengan lipid dan protein pada membran sel lensa, memicu ikatan lipid dan protein, penggumpalan protein, serta peningkatan jumlah protein yang tidak larut dalam air. Hal ini mengakibatkan penurunan kejernihan lensa dan memicu terbentuknya katarak. Kerusakan ini dapat disebabkan oleh hidrasi berlebihan pada lensa, denaturasi protein lensa, serta gangguan pada metabolisme normal lensa yang biasanya terjadi seiring bertambahnya usia. [4].

Salah satu metode data mining adalah klasifikasi, yang memungkinkan pengelompokan data. Beberapa peneliti pernah melakukan penelitian tentang klasifikasi penyakit mata. Penelitian pertama, Penelitian oleh Rizka Nurlizah dan kawan-kawan yang berjudul Klasifikasi Penyakit Katarak Pada Mata Manusia Menggunakan Metode Convolutional Neural Network. Penelitian ini berfokus pada pembagian jenis mata katarak untuk menentukan apakah mata pasien memiliki penyakit katarak atau tidak, sehingga dokter dapat menangani pasien dengan cepat. Pengolahan gambar akan dilakukan dalam format JPG/JPEG yang lebih umum, dan metode CNN (Convolutional Neural Network) digunakan di sini [5].

Penelitian kedua yang dilakukan oleh Anita Sindar dan rekan-rekannya, berjudul Diagnosa Penyakit Mata Berdasarkan Citra Ocular Disease Intelligent Recognition (ODIR) dengan Gabor Filter Klasifikasi Levenberg-Marquardt, berfokus pada diagnosis penyakit retina berdasarkan penglihatan manusia. Studi ini menggunakan data latih dari database ODIR untuk menganalisis struktur visual retina. [6].

Penelitian ketiga yang dilakukan oleh Ayu Evilia Suwanda dan tim, dengan judul Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Retina Menggunakan Dimensi Fraktal Box Counting dan Fuzzy K-Means, bertujuan untuk mengevaluasi apakah metode penghitungan dimensi fraktal pada citra fundus retina dapat digunakan untuk mendeteksi penyakit mata melalui klasifikasi fuzzy k-means. Dari hasil penelitian ini, diperoleh akurasi sebesar 76%, sehingga disimpulkan bahwa teknik dimensi fraktal box counting pada citra fundus retina dapat dimanfaatkan untuk mengidentifikasi jenis penyakit mata [7].

Penggerak mekanis memudahkan dalam membedakan dan mengklasifikasikan katarak melalui penanganan gambar tingkat lanjut. Dalam tahap terakhir ini, pencipta menggunakan proses Jaringan neural konvolusi berbasis web untuk membentuk aplikasi pembelajaran mesin yang dapat memisahkan mata penderita katarak dari mata normal. Informasi tersebut berasal dari lokasi ilmu informasi bernama Kaggle. Dataset yang digunakan berasal dari gambar lanjutan yang berisi dua peningkatan pembelajaran, yaitu kursus katarak dan pembelajaran biasa. Hasil terakhir yang diantisipasi adalah model yang dibuat mampu membedakan mata tipikal dan mata katarak.

2. TUJUAN PENELITIAN

Penelitian ini bertujuan untuk mendeteksi dini penyakit katarak pada mata dengan menggunakan metode Convolutional Neural Network (CNN) berbasis pengolahan citra dan *machine learning*. Penelitian ini diharapkan dapat membangun model prediktif yang mampu membedakan mata normal dari mata yang mengalami katarak dengan akurasi yang tinggi. Dengan demikian, model ini dapat membantu mempercepat diagnosis katarak sehingga intervensi medis dapat dilakukan lebih awal dan mengurangi risiko kebutaan di Indonesia.

3. TINJAUAN PUSTAKA

Tinjauan ini akan membantu menempatkan penelitian ini dalam konteks yang tepat, memastikan pendekatan yang diambil berdasarkan metodologi yang terbukti, serta memperkuat argumen tentang efektivitas CNN dalam mendeteksi katarak secara otomatis dan dini.

3.1 Katarak

Katarak adalah kondisi di mana lensa mata menjadi keruh sebagian atau seluruhnya, sehingga cahaya sulit menembusnya dan menyebabkan penglihatan tampak kabur, seolah-olah tertutup kabut. Pada mata yang sehat, lensa tetap jernih tanpa ada tanda-tanda kekeruhan. Ada berbagai jenis katarak, namun penelitian ini berfokus pada katarak senilis, yang terjadi pada orang lanjut usia. Katarak yang muncul pada orang tua disebut sebagai katarak senilis [8]. Kekeruhan pada lensa mata merupakan gejala utama katarak, yang umumnya terjadi pada individu di atas usia 50 tahun. Namun, kondisi ini juga bisa muncul akibat kelainan bawaan atau penyakit mata kronis. Katarak dapat dipicu oleh berbagai penyakit mata, seperti glaukoma, uveitis, ablasi, dan retinitis pigmentosa, serta dapat berhubungan dengan masalah intraokular lainnya. Meskipun katarak paling sering dikaitkan dengan penuaan, kondisi ini juga bisa terjadi pada orang yang lebih muda, termasuk mereka yang berusia di bawah lima puluh tahun. Katarak bisa berkembang di salah satu atau kedua mata, dan dalam satu mata dapat terdapat lebih dari satu jenis katarak [9].

3.2 Deep Learning Deep

Salah satu komponen kecerdasan buatan yang pertama kali terinspirasi oleh struktur otak manusia adalah deep learning. Teknologi ini memungkinkan komputer untuk "belajar" dari data dengan tingkat akurasi yang lebih tinggi. Deep learning memiliki kemampuan luar biasa dalam mengenali dan mengidentifikasi berbagai jenis data,

seperti teks, gambar, dan video, dengan presisi yang sangat tinggi [10]. Model deep learning mampu mempelajari pola komputasinya sendiri dengan cara kerja yang menyerupai otak manusia. Deep learning dirancang untuk menganalisis data seperti bagaimana otak manusia bekerja ketika membuat keputusan. Teknologi ini menggunakan algoritma artificial neural network (ANN), yang terinspirasi oleh jaringan saraf biologis di otak manusia [11].

3.3 Convolutional Neural Network (CNN)

CNN merupakan pengembangan dari Multilayer Perceptron (MLP) yang dirancang khusus untuk menangani data dalam bentuk dua dimensi. Karena memiliki lapisan jaringan yang dalam, CNN dikategorikan sebagai bagian dari Deep Neural Network dan sering digunakan untuk pemrosesan data citra. CNN pertama kali dikembangkan oleh Kunihiko Fukushima, seorang peneliti dari Laboratorium Penelitian Sains Broadcasting NHK di Kinuta, Setagaya, Tokyo, Jepang, dengan nama NeoCognitron [12]. Convolutional Neural Network (CNN) adalah metode deep learning yang mampu melakukan proses pembelajaran mandiri untuk pengenalan objek. Teknik ini dapat melakukan ekstraksi dan klasifikasi objek, serta diterapkan pada gambar dengan resolusi tinggi untuk menghasilkan model distribusi nonparametrik [13]. Pada dasarnya, CNN menggunakan metode konvolusi dengan cara menggeser filter (saluran) berukuran tertentu pada sebuah gambar. Dari proses ini, komputer dapat memperoleh informasi penting dengan memanfaatkan hasil pemrosesan dari bagian gambar yang dilewati filter. CNN merupakan arsitektur jaringan saraf tiruan yang sangat efisien untuk menangani tugas-tugas yang melibatkan data gambar dan video. Dengan memanfaatkan lapisan konvolusi untuk mengekstrak fitur lokal serta lapisan yang mengurangi dimensi data, CNN mampu mempelajari representasi fitur yang sangat kompleks dan melakukan klasifikasi dengan tingkat akurasi yang tinggi.

3.4 Citra Digital

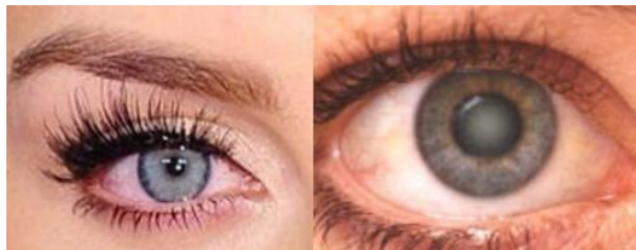
Istilah lain untuk menyebut gambar sebagai elemen multimedia yang krusial dalam menyampaikan informasi visual adalah citra. Gambar yang kaya akan informasi memiliki karakteristik unik yang tidak dimiliki oleh data teks. Secara harfiah, gambar, atau visual, adalah representasi dua dimensi. Dari sudut pandang matematis, gambar dapat dipahami sebagai fungsi kontinu dari intensitas cahaya pada bidang dua dimensi. Ketika objek terkena cahaya dari sumber, sebagian dari cahaya tersebut dipantulkan kembali oleh objek [14].

4. METODE PENELITIAN

Penelitian ini bertujuan untuk mengembangkan model yang dapat mengklasifikasikan katarak pada mata dengan menggunakan metode Convolutional Neural Network (CNN). Untuk mencapai tujuan tersebut, sejumlah langkah strategis telah dilakukan, antara lain :

4.1 Pengumpulan Data

Data yang digunakan adalah citra mata yang diperoleh dari berbagai sumber, dataset yang tersedia untuk umum. Dataset ini berisi gambar mata yang diklasifikasikan menjadi beberapa kategori, seperti mata normal dan mata katarak. Setiap gambar dilabeli sesuai dengan kondisi mata untuk memudahkan proses pelatihan model. Data dipakai untuk melatih di dapatkan di web. Ada 2112 gambar yang terdiri dari mata normal dan mata katarak. Contoh masing-masing jenis mata ditunjukkan pada Gambar 1.



Gambar 1. Mata Normal dan Mata Katarak Pada Manusia

4.2 Data Preprocessing

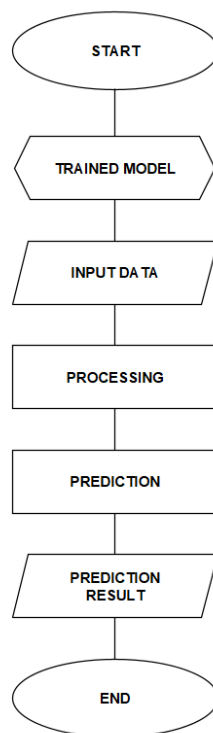
Tahap persiapan data, atau preprocessing data, adalah proses mengolah data awal agar siap digunakan pada model klasifikasi [15]. Sebelum digunakan, dilakukan pra-pemrosesan data. Preprocessing data digunakan dalam pembelajaran mesin untuk memastikan bahwa data yang digunakan berkualitas baik, yang kemudian dapat digunakan dalam tampilan pembelajaran mesin yang akan dibangun. dataset yang digunakan dibagi dua bagian, yaitu latih dan tes. Oleh karena itu, pengorganisasian pra-pemrosesan informasi utama dilakukan untuk menormalkan penentuan gambar-gambar yang digunakan, atau mengubah perkiraannya, sehingga demonstrasi

yang dibangun dapat dibuat secara akurat. Gambar yang diambil tentang katarak dan mata biasa adalah gambar dengan ukuran berbeda-beda lalu dipotong agar ringan. Ukuran penyajian ubah jadi 150 x 150 piksel.

4.3 Pelatihan Model

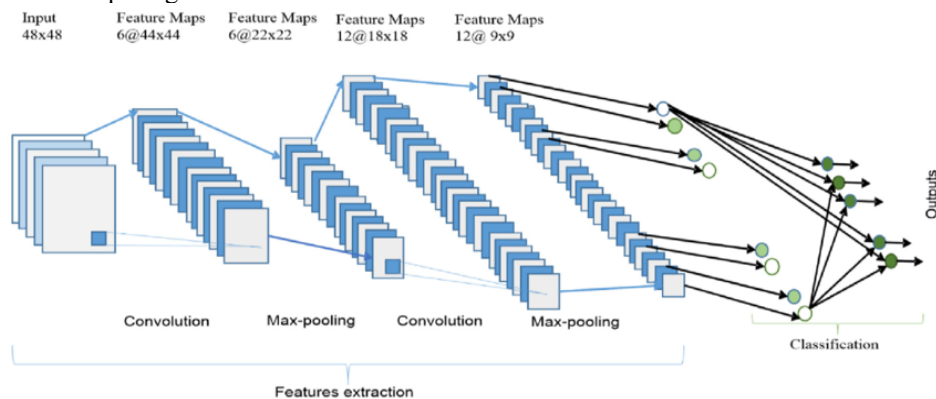
Proses pelatihan dilakukan dengan menggunakan *dataset* yang telah diproses terlebih dahulu untuk meningkatkan performa model. Data kemudian dipecah menjadi set latihan dan set tes dengan proporsi tertentu, misalnya 80% untuk pelatihan dan 20% untuk tes, guna mengevaluasi performa model. Model dilatih menggunakan algoritma optimasi Adam, dengan parameter *learning rate* yang telah ditentukan, untuk mengontrol seberapa besar langkah pembaruan bobot selama pelatihan.

Fungsi *loss* yang digunakan adalah *sparse categorical crossentropy*, yang sesuai untuk tugas klasifikasi multikelas dengan label integer. Setelah menjalankan proses pelatihan selama sejumlah epoch dan memantau metrik kinerja seperti akurasi, diperoleh model jaringan saraf tiruan yang layak digunakan untuk melakukan prediksi pada data baru. Tahapan-tahapan ini secara umum mencakup menyiapkan model latih, memasukkan data latih, memproses data, memprediksi data dan menghasilkan data prediksi yang dapat dijelaskan melalui diagram alur sistem pada Gambar 2 berikut.



Gambar 2. Alur Sistem

Langkah utama adalah pemrosesan awal data. Dalam hal ini, ubah ukuran gambar kumpulan data menjadi 150 x 150 piksel. Data tersebut pisah menjadi dua, yaitu latih dan tes, dengan perbandingan 80:20. Pada bagian pelatihan dan tes terdapat 1038 gambar mata katarak, dan pada bagian latihan dan tes terdapat 1074 gambar mata normal. Selanjutnya informasi persiapan yang terkumpul disusun menggunakan tayangan CNN dalam beberapa tahap, seperti terlihat pada gambar 3.



Gambar 3. Layer Model CNN

Sebelum model dapat melakukan klasifikasi, persiapan informasi harus melalui beberapa lapisan, seperti terlihat pada gambar 3. Pada pengorganisasian pertama, persiapan informasi ditangani melalui lapisan *convolution*, lapisan *pooling*, dan lapisan *flatten* untuk dibuat.

Tabel 1. Lapisan Model CNN dengan Keras Sequential

Layer (type)	Output Shape	Param #
inception_v3 (Functional)	(None, 3, 3, 2048)	21,802,784
conv2d_91 (Conv2D)	(None, 1, 1, 32)	589,856
max_pooling2d_4 (MaxPooling2D)	(None, 1, 1, 32)	0
dropout (Dropout)	(None, 1, 1, 32)	0
batch_normalization_94 (BatchNormalization)	(None, 1, 1, 32)	128
max_pooling2d	(None, 1, 1, 32)	0
dropout_1 (Dropout)	(None, 1, 1, 32)	0
flatten (Flatten)	(None, 32)	0
dense (Dense)	(None, 2)	66

Rumus evaluasi :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Keterangan :

TP = jumlah kasus di mana model berhasil mendeteksi dengan benar sesuatu yang positif

TN = jumlah kasus di mana model berhasil mendeteksi dengan benar sesuatu yang negatif

FN = jumlah kasus di mana model salah mendeteksi positif ketika seharusnya negatif

Accuracy = Mengukur seberapa sering model membuat prediksi yang benar, baik untuk kasus positif maupun negatif. menunjukkan proporsi dari total prediksi yang benar.

Precision = Mengukur berapa banyak dari prediksi positif yang benar-benar positif. berarti model jarang salah dalam memprediksi positif.

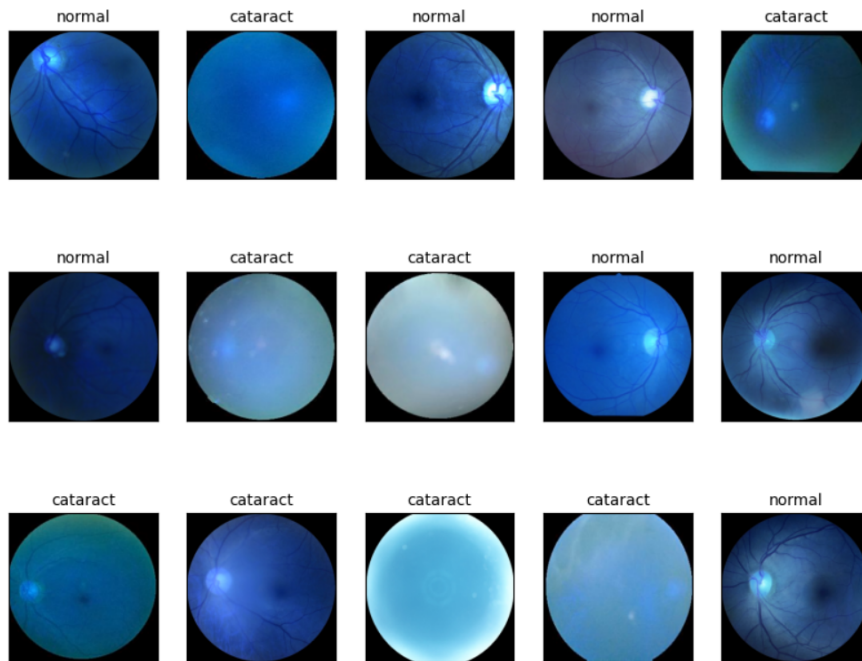
Recall = Mengukur seberapa banyak dari kasus positif yang sebenarnya berhasil dideteksi oleh model. Recall menunjukkan sensitivitas model terhadap kasus positif. Semakin tinggi recall, semakin sedikit kasus positif yang terlewat oleh model.

F1-score = Merupakan rata-rata harmonik antara presisi dan recall. berguna saat ada ketidakseimbangan antar data positif dan negatif, karena ini memberikan keseimbangan antara presisi dan recall. Nilai yang lebih tinggi menunjukkan kinerja model yang lebih baik dalam hal mendeteksi kelas positif tanpa terlalu banyak kesalahan.

4. HASIL DAN PEMBAHASAN

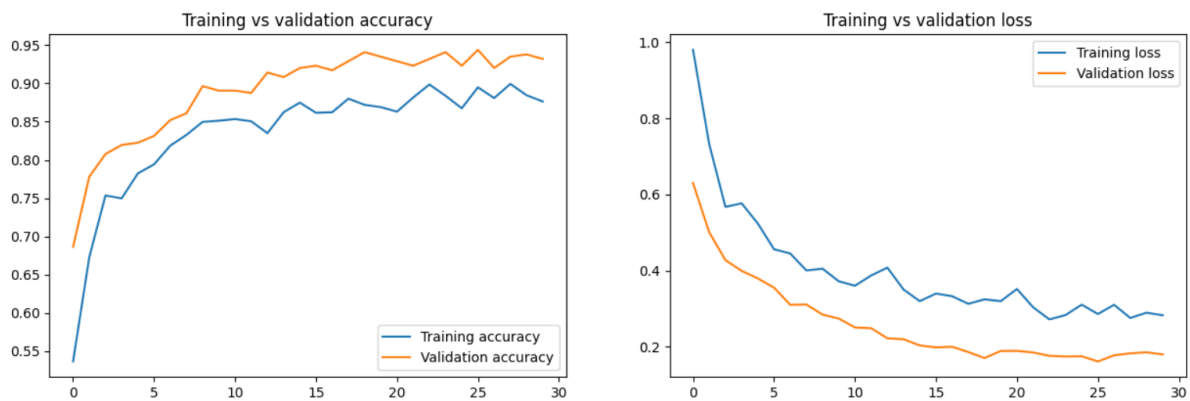
Penggunaan citra retina dalam diagnosis katarak merupakan salah satu contoh dari aplikasi teknologi ini. Dengan bantuan algoritma yang canggih dan teknik pengolahan citra seperti yang digunakan dalam Convolutional Neural Network (CNN), sistem otomatis dapat dilatih untuk mengenali tanda-tanda katarak pada gambar retina. Mata yang terkena katarak akan menunjukkan area keruh pada lensa, yang biasanya terlihat sebagai lingkaran berwarna putih atau abu-abu di tengah gambar. Intensitas cahaya yang memasuki retina dapat terlihat lebih redup atau menyebar. Dengan melihat warna dan kejelasan lensa di tengah gambar, kita bisa membedakan antara mata

yang normal dan yang terkena katarak. Pada citra yang normal, area tengah tampak lebih jernih dan pada citra yang terkena katarak menunjukkan area tengah terlihat keruh atau buram.



Gambar 4 Hasil CNN Dataset Mata Normal dan Katarak

Gambar di atas menampilkan 15 sampel citra retina yang dipilih secara acak, dengan label normal atau cataract di bagian atas setiap gambar. Citra tersebut diperoleh dari dataset, di mana masing-masing gambar menunjukkan kondisi retina, baik normal maupun yang terkena katarak. Citra retina dihasilkan menggunakan fungsi OpenCV (cv2.imread) untuk membaca gambar dari filepaths, dan setiap gambar diberikan label sesuai kategori yang telah ditentukan dalam dataset.



Gambar 5 Perbandingan Antara Accuracy dan Loss dari Data Training dan Data Validasi

Grafik di sebelah kiri menunjukkan *accuracy* (akurasi) model selama proses training dan validasi. Awalnya, baik *training accuracy* maupun *validation accuracy* meningkat seiring bertambahnya epoch, dengan *validation accuracy* mencapai nilai yang lebih tinggi dan lebih stabil di sekitar epoch ke-20.

Grafik di sebelah kanan menunjukkan *loss* (nilai kesalahan) model selama training dan validasi. Kedua grafik *loss* terus menurun, menunjukkan bahwa model semakin baik dalam memprediksi data. Namun, *validation loss* memiliki nilai yang lebih rendah dan menurun lebih cepat daripada *training loss*, yang menandakan generalisasi model yang baik terhadap data validasi.

Dalam penelitian ini, model dibangun dengan epoch 30 dan ukuran batch 32. Setelah itu, perhitungan kerugian kerja dan nilai ketepatan digunakan untuk menentukan apakah ekspektasi yang terjadi baik atau buruk. Untuk perbandingan dapat di lihat pada Gambar 6.

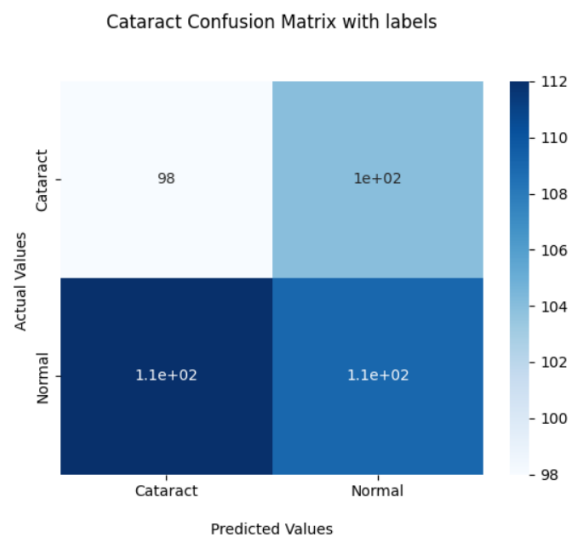
Tabel 2 Hasil klasifikasi

	Precision	Recall	F1-score	Support
Cataract	0.46667	0.48515	0.48515	202
Normal	0.51174	0.49321	0.50230	221
Accuracy			0.48936	423
macro avg	0.48920	0.48918	0.48902	423
weighted avg	0.49021	0.48936	0.48961	423

Confusion Matrix nya adalah : $\begin{bmatrix} 98 & 104 \\ 112 & 109 \end{bmatrix}$

Sesuai dengan tabel 2, maka nilai presisi yang dihasilkan untuk cataract sebesar 46,67%, nilai recall cataract sebesar 48,52%, nilai f1-score cataract sebesar 0,47573 dan nilai support cataract sebesar 202 sampel yang digunakan. Dengan demikian, model ini mempunyai kemampuan untuk mengidentifikasi katarak.

Dari 202 sampel pada baris pertama yang sebenarnya cataract, 98 diprediksi benar sebagai cataract, dan 104 salah diprediksi sebagai normal. Dari 221 pada baris kedua sampel yang sebenarnya normal, 109 diprediksi benar sebagai normal, dan 112 salah diprediksi sebagai cataract.



Gambar 6 Nilai Prediksi Katarak Confusion Matrix dan Label

Untuk menghitung evaluasi kinerja model yang dibangun, langkah selanjutnya adalah menghitung matriks konfusi. Dapat melihat pada Gambar 6.

5. DISKUSI

Penelitian ini memiliki beberapa keterbatasan, di antaranya tidak mencantumkan penjelasan mengenai gejala serta penyebab terjadinya katarak, yang merupakan informasi penting dalam memahami latar belakang patologi katarak. Selain itu, hasil deteksi yang diperoleh hanya bersifat umum tanpa pemisahan rinci berdasarkan jenis katarak, seperti katarak senilis, traumatik, juvenil, maupun subkapsularis posterior, yang seharusnya dapat memperkaya analisis deteksi. Terakhir, penelitian ini belum dilengkapi dengan uji kredibilitas sistem untuk memastikan validitas dan reabilitas deteksi dalam berbagai skenario, yang penting untuk meningkatkan keandalan model secara keseluruhan.

6. KESIMPULAN

Penelitian yang dilakukan terhadap klasifikasi katarak menggunakan metodologi Convolutional Neural Network (CNN) telah menghasilkan performa yang dapat diukur. Berdasarkan hasil klasifikasi, penelitian ini mampu membedakan antara mata yang terkena katarak dan mata normal. Prediksi menunjukkan bahwa akurasi

model mencapai 48.94%, dengan nilai precision sebesar 46.67% untuk kelas katarak dan 51.17% untuk kelas normal. Nilai recall yang dicapai adalah 48.52% untuk katarak dan 49.32% untuk normal, sementara nilai F1-score masing-masing adalah 47.57% untuk katarak dan 50.23% untuk normal.

Selain itu, seperti yang ditunjukkan oleh grafik perubahan presisi (precision) dan kerugian (loss) pada setiap epoch, pengaturan penanganan model dilakukan dengan baik, meskipun ada beberapa potensi peningkatan terutama dalam generalisasi model. Akurasi selama proses pelatihan dan validasi menunjukkan bahwa meskipun ada peningkatan performa selama proses pelatihan, hasil pada data validasi menunjukkan adanya ketidakseimbangan yang perlu diperhatikan untuk mencapai performa yang lebih baik.

DAFTAR PUSTAKA

- [1] J. Kecerdasan, T. Informasi, D. H. Firdaus, B. Imran, L. D. Bakti, and E. Suryadi, "Klasifikasi Penyakit Katarak Pada Mata Menggunakan Metode Convolutional Neural Network (Cnn) Berbasis Web Web-Based Classification of Cataract in the Eyes Using Convolutional Neural Network (Cnn) Method," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 1, no. 3, pp. 18–26, 2022.
- [2] A. U. Detty, I. Artini, and V. R. Yulian, "Karakteristik Faktor Risiko Penderita Katarak," *J. Ilm. Kesehat. Sandi Husada*, vol. 10, no. 1, pp. 12–17, 2021, doi: 10.35816/jiskh.v10i1.494.
- [3] F. Ramadhani, A. Satria, and S. Salamah, "Implementasi Algoritma Convolutional Neural Network dalam Mengidentifikasi Dini Penyakit pada Mata Katarak," *sudo J. Tek. Inform.*, vol. 2, no. 4, pp. 167–175, 2023, doi: 10.56211/sudo.v2i4.408.
- [4] G. A. Wiguna, R. Fardela, and J. B. Selly, "Klasifikasi Tingkat Maturitas Katarak Berbasis Citra Digital Berdasarkan Jangkauan (Range) Nilai Histogram," *J. Saintek Lahan Kering*, vol. 2, no. 2, pp. 54–57, 2019, [Online]. Available: <https://doi.org/10.32938/slk.v2i2.869>
- [5] E. Ayuningrum, A. E. Minarno, and G. W. Wicaksono, "Klasifikasi Penyakit Katarak Pada Mata Manusia Menggunakan Metode Convolutional Neural Network," *J. Repos.*, vol. 4, no. 4, pp. 491–496, 2024, doi: 10.22219/repositor.v4i4.32285.
- [6] A. S. Sinaga, F. Ginting, and S.- Ramen, "Diagnosa Penyakit Mata Berdasarkan Citra Ocular Disease Intelligent Recognition (ODIR) Dengan Gabor Filter Klasifikasi Levenberg-Marquardt," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 22, no. 2, p. 432, 2023, doi: 10.53513/jis.v22i2.8787.
- [7] D. Juniati and A. E. Suwanda, "Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Retina Menggunakan Dimensi Fraktal Box Counting Dan Fuzzy K-Means," *Prox. J. Penelit. Mat. dan Pendidik. Mat.*, vol. 5, no. 1, pp. 10–18, 2022, doi: 10.30605/proximal.v5i1.1623.
- [8] K. Representasi and M. Siswa, "3 1,2,3," vol. 7, no. 2, pp. 5054–5061, 2018.
- [9] R. Z. Ithram, R. D. Atmaja, and I. Widjayanto, "Deteksi Dan Klasifikasi Stadium Katarak Senilis Berdasarkan Citra Mata Menggunakan Metode Support Vector Machine (Svm) Vector Machine (Svm)," *e-Proceeding Eng.*, vol. 5, no. 2, pp. 2023–2030, 2018.
- [10] Royan Hisyam Rafliansyah, Basuki Rahmat, and Chrystia Aji Putra, "Klasifikasi Suara Instrumen Musik Tiup Menggunakan Metode Convolutional Neural Network," *Merkurius J. Ris. Sist. Inf. dan Tek. Inform.*, vol. 2, no. 4, pp. 01–09, 2024, doi: 10.61132/mercurius.v2i4.119.
- [11] A. Peryanto, A. Yudhana, and R. Umar, "Rancang Bangun Klasifikasi Citra Dengan Teknologi Deep Learning Berbasis Metode Convolutional Neural Network," *Format J. Ilm. Tek. Inform.*, vol. 8, no. 2, p. 138, 2020, doi: 10.22441/format.2019.v8.i2.007.
- [12] Verdy and Ery Hartati, "Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network Model Resnet-50," *J. Rekayasa Sist. Inf. dan Teknol.*, vol. 1, no. 3, pp. 199–206, 2024, doi: 10.59407/jrsit.v1i3.529.
- [13] K. Wahyuddin, D. Wahiddin, and D. . Kusumaningrum, "Door Security Face Detection System Using Arduino-based Convolutional Neural Network (CNN) Method," *Sci. Student J. Information, Technol. Sci.*, vol. 4, no. 1, pp. 15–23, 2023.
- [14] I. P. E. Sutariawan, G. R. Dantes, and K. Y. Ernanda Aryanto, "Segmentasi Mata Katarak Pada Citra Medis Menggunakan Metode Operasi Morfologi," *J. Ilmu Komput. Indones.*, vol. 3, no. 1, pp. 23–31, 2019, doi: 10.23887/jik.v3i1.2750.
- [15] Mohammad Haydir Awaludin Waskito, Andreas Nugroho Sihananto, and Achmad Junaidi, "Klasifikasi Penyakit Kronis Melalui Mata Menggunakan Algoritma Convolutional Neural Network Dengan Model

MobileNet-V3,” *Uranus J. Ilm. Tek. Elektro, Sains dan Inform.*, vol. 2, no. 2, pp. 48–60, 2024, doi: 10.61132/uranus.v2i2.120.

SENASTIKA Universitas Malikussaleh

ANALISIS HYBRID DECISION SUPPORT SYSTEM DALAM PEMILIHAN LOKASI USAHA (STUDI KASUS: KOTA IDI RAYEUK KABUPATEN ACEH TIMUR)

Husnal Almaiyah¹, Munirul Ula², Sujacka Retno³

¹²³Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

¹husnaluun@gmail.com

Abstract

Choosing the right business location is one of the crucial factors that can influence the success of a business. In this research, a Hybrid Decision Support System (DSS) analysis was carried out for business location selection using the Analytical Hierarchy Process (AHP) method and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The criteria used in selecting this location include rental price, shop equipment, parking space, number of businesses around the location, target market, security, crowd level, and distance of the location to the city center. This research examines three alternative locations, namely Kampung Jawa, Keude Blang, and Tanah Anou. The analysis results show that Tanah Anou has the highest score in the ranking, making it the most ideal location choice based on predetermined criteria. Thus, the use of the AHP and TOPSIS methods in this DSS has proven to be effective in assisting decision making in selecting business locations.

Keywords: Hybrid DSS, AHP, TOPSIS, Business Location Selection, Decision Support System.

Abstrak

Pemilihan lokasi usaha yang tepat merupakan salah satu faktor krusial yang dapat mempengaruhi keberhasilan sebuah bisnis. Dalam penelitian ini, dilakukan analisis *Hybrid Decision Support System (DSS)* untuk pemilihan lokasi usaha dengan menggunakan metode *Analytical Hierarchy Process (AHP)* dan *Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)*. Kriteria yang digunakan dalam pemilihan lokasi ini meliputi harga sewa, kelengkapan toko, lahan parkir, jumlah usaha di sekitar lokasi, target pasar, keamanan, tingkat keramaian, dan jarak lokasi dengan pusat kota. Penelitian ini mengkaji tiga alternatif lokasi, yaitu Kampung Jawa, Tanah Anou dan Keude Blang. Hasil analisis menunjukkan bahwa Keude Blang memiliki nilai tertinggi dalam perankingan, menjadikannya sebagai pilihan lokasi yang paling ideal berdasarkan kriteria yang telah ditentukan. Dengan demikian, penggunaan metode AHP dan TOPSIS dalam DSS ini untuk membantu pengambilan keputusan dalam pemilihan lokasi usaha.

Kata kunci: Hybrid DSS, AHP, TOPSIS, Pemilihan Lokasi Usaha, Sistem Pengambilan Keputusan.

1. PENDAHULUAN

Di era globalisasi ini perkembangan teknologi perangkat lunak melaju dengan sangat pesat, ini dapat ditinjau dari munculnya berbagai aplikasi. Dari pesatnya perkembangan ini perluasan pemanfaatan komputer yang semula dimanfaatkan sedikit orang, kini menjadi dimanfaatkan banyak orang. Hal ini mengakibatkan perubahan pada berbagai bidang kehidupan seperti halnya dalam proses pengambilan sebuah keputusan, terkadang keputusan yang dibuat oleh seseorang atau sekelompok kurang akurat dalam penilaiannya, Dalam penelitian ini bahasannya yaitu tentang pemilihan lokasi usaha yang hanya mencakup penyewaan ruko tempat usaha dan tidak memasukkan lokasi seperti tanah kosong atau rumah biasa. Menurut hasil observasi yang telah dilakukan bahwa dalam proses pemilihan lokasi usaha masih bersifat subyektif, sehingga keputusan yang dibuat kurang sesuai dengan kondisi [1].

Idi Rayeuk adalah salah satu kecamatan di Kabupaten Aceh Timur, Provinsi Aceh, Indonesia, dan sebuah kota yang sedang mengalami perkembangan signifikan dari berbagai hal baik segi ekonomi dan kebudayaan sehingga menjadi sebuah tempat tujuan untuk melakukan usaha maupun tujuan sekedar tempat untuk rekreasi. Hal tersebut menyebabkan menjamurnya usaha – usaha kecil menengah.

Memilih lokasi yang strategis sebagai tempat untuk membuka usaha sangatlah tidak mudah, karena membutuhkan lokasi yang tepat dan sesuai agar usaha yang dijalankan dapat diterima oleh konsumen dan usaha nya tidak cepat gulung tikar. Namun, karena perkembangan idi rayeuk yang sangat pesat, pencarian tempat oleh pemilik usaha menjadi sulit karena titik atau lokasi yang bagus tersebut tidak terlalu banyak

sehingga pencarian secara manual pasti akan membutuhkan waktu yang lama dan membuang sumber daya yang lumayan juga menjadi tidak efektif, apalagi di zaman yang serba digital sekarang[2].

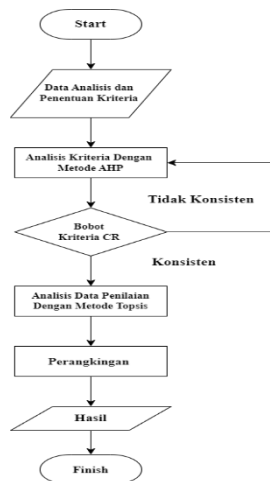
Berdasarkan masalah diatas maka diperlukan suatu sistem yang bisa mengkalkulasikan segala kriteria yang mendukung dalam pemilihan sebuah keputusan untuk memilih titik atau lokasi untuk mendirikan sebuah usaha yang bagus menggunakan metode AHP (Analytical Hierarchy Process) dan TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) sehingga dapat membantu dan mempermudah proses pengambilan keputusan [3].

Dalam pengembangannya metode AHP dapat dikombinasikan dengan tujuan meminimalkan kesalahan proses pembobotan berdasarkan kriteria Pembobotan yang didapat akan dimaksimalkan metode TOPSIS untuk melakukan perhitungan guna mendapatkan perankingan berdasarkan keluaran. Metode TOPSIS memberikan hasil keluaran dengan bentuk jarak solusi ideal dari kriteria. Jarak yang dihasilkan akan dianalisis dalam gambaran hasil perankingan untuk dijadikan solusi dalam penyelesaian sebuah masalah. Dengan penjelasan tersebut, metode AHP dan Topsis dapat memaksimalkan proses analisis guna memberikan hasil solusi yang lebih baik [4].

2. METODELOGI PENELITIAN

Penelitian ini mengadopsi pendekatan dengan model kualitatif dengan menyajikan perhitungan matematis dalam menyelesaikan sebuah masalah. Parameter pendekatan yang digunakan didasari pada konsep serta metode yang mengadopsi perhitungan secara matematis dalam melakukan analisis [4].

Adapun skema sistem tersebut terdapat pada gambar dibawah ini:



Gambar 1. Skema Sistem

Gambar 1. Menjelaskan Skema Penelitian yang dimulai dari proses analisa data, pembentukan kriteria, analisis dengan metode AHP, analisis dengan metode TOPSIS untuk mendapatkan hasil keputusan berupa perankingan.

Pada proses analisis AHP hasil yang diharapkan mampu memberikan nilai bobot yang konsisten, nilai konsistensi yang disebut dengan nilai rasio consistency (CR) yaitu $< 0,1$ yang merupakan nilai ambang batas (threshold) yang sudah ditetapkan oleh penemu metode ini. Jika hasil nilai $CR > 0,1$ maka nilai kriteria perbandingan berpasangan yang sudah diuji hasilnya tidak konsisten dan diulang kembali pemilihan nilai perbandingan berpasangan pada langkah ke 3. Jika nilai $CR < 0,1$ maka hasilnya konsisten dimana nilai bobotnya dapat digunakan untuk langkah selanjutnya yaitu analisis data dengan metode TOPSIS. Pada analisis metode TOPSIS, nilai bobot yang sudah konsisten dari metode AHP akan dikalkulasikan dalam menentukan matriks keputusan ternormalisasi terbobot pada langkah ke 3, dan selanjutnya nilai hasil keputusan disajikan dalam bentuk perankingan.

Adapun kriteria yang digunakan dapat dilihat pada Tabel 1.

Kriteria	Keterangan	Jenis Atribut
C1	Harga sewa	Cost
C2	Kelengkapan Toko	Benefit
C3	Lahan Parkir	Benefit
C4	Jumlah Usaha Disekitar Lokasi	Cost
C5	Target Pasar	Benefit
C6	Keamanan	Benefit
C7	Tingkat Keramaian	Benefit
C8	Jarak Lokasi Dengan Pusat kota	Benefit

Tabel 1. Kriteria

Tabel 1 Menjelaskan bahwa kriteria yang digunakan dalam penilaian diinisialisasi dengan kode C1, C2, C3, C4, C5, C6, C7 dan C8. Adapun gambaran metode AHP dapat lihat berdasarkan Langkah-langkah berikut [4]:

1. Mendefinisikan masalah, menentukan tujuan, kriteria, sub kriteria dan alternatif solusi.
2. Membuat struktur hirarki dari atas ke bawah yaitu tujuan, kriteria, sub kriteria dan alternatif solusi.
3. Membuat matriks perbandingan berpasangan (pairwise comparison) kriteria. Perbandingan dilakukan berdasarkan “judgment” dari pengambil keputusan dengan menilai tingkat kepentingan suatu elemen dibandingkan elemen lainnya.
4. Nilai kriteria perbandingan berdasarkan pada Tabel 2.

Tingkat Kepentingan	Definisi	Keterangan
1	Sama Pentingnya	Kedua elemen mempunyai pengaruh yang sama
3	Sedikit Lebih Penting	1 level lebih penting dibandingkan kriteria lainnya
5	Lebih Penting	2 level lebih penting dibandingkan kriteria lainnya
7	Sangat Penting	3 level lebih penting dibandingkan kriteria lainnya
9	Mutlak Lebih Penting	4 level lebih penting dibandingkan kriteria lainnya
2, 4, 6, 8	Nilai Tengah	Diberikan bila terdapat keraguan penilaian di antara dua tingkat kepentingan yang berdekatan.

Tabel 2. Skala Penilaian Perbandingan Berpasangan

5. Menghitung matriks bobot nilai antar kriteria dan prioritas.
6. Menghitung matrik penjumlahan setiap baris atau Menghitung vektor eigen dari setiap matriks perbandingan berpasangan.
7. Menghitung nilai rasio konsistensi kriteria. Jika nilai rasio consistency (CR) ≤ 0,1 maka penilaian yang dilakukan adalah konsisten, dengan menggunakan Persamaan 1 & 2 [5] :

$$CI = \frac{\lambda_{maks} - n}{n - 1} \tag{2.1}$$

$$CR = \frac{CI}{IR} \tag{2.2}$$

Menentukan nilai indeks random (IR) dapat dilihat pada Tabel 3.

Ordo Matrik	R1	Ordo Matrik	R1	Ordo Matrik	R1
1	0	6	1,24	11	1,51
2	0	7	1,32	12	1,48
3	0,58	8	1,41	13	1,56
4	0,9	9	1,45	14	1,57
5	1,12	10	1,49	15	1,57

Tabel 3. Nilai Indeks Random

Tabel 3 dapat dilihat bahwa nilai indeks random yang digunakan dalam perhitungan metode AHP terdiri dari Ordo Matriks dan Nilai Indeks Random. Tahapan selanjutnya adalah melakukan perhitungan dengan menggunakan metode Topsis.

Salah satu metode pengambilan keputusan multi kriteria adalah *Technique For Order Preference By Similarity To Ideal Solution* (TOPSIS). Dalam TOPSIS terdapat prinsip bahwa alternatif yang terpilih

harus mempunyai jarak terdekat dari solusi ideal positif dan jarak terpanjang (terjauh) dari solusi ideal negatif dari sudut pandang geometris [6]

Adapun Langkah-langkah yang akan dilakukan dapat dilihat sebagai berikut :

1. TOPSIS dimulai dengan membangun sebuah matriks keputusan. Matriks keputusan X mengacu terhadap m alternatif yang akan dievaluasi berdasarkan n kriteria. Matriks keputusan X dapat dilihat pada Persamaan 3.

$$X = a_1 : a_m [x_{11} \dots x_{1n} : \dots : a_{m1} \dots x_{mn}] \quad (2.3)$$

Persamaan 4 menjelaskan bahwa a_i ($i = 1, 2, 3, \dots, m$) adalah alternatif-alternatif yang mungkin, x_j ($j = 1, 2, 3, \dots, n$) adalah atribut dimana performansi alternatif diukur, x_{ij} adalah performansi alternatif a_i dengan acuan atribut x_j .

2. Membuat matriks bobot keputusan yang ternormalisasi. TOPSIS membutuhkan rating kinerja setiap alternatif A_i pada setiap bobot r_{ij} yang ternormalisasi seperti pada Persamaan 4.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \quad (2.4)$$

Dengan $i = 1, 2, 3, \dots, m$; dan $j = 1, 2, 3, \dots, n$; dimana r_{ij} adalah elemen dari matriks keputusan yang ternormalisasi R. x_{ij} adalah elemen matriks dari keputusan X.

3. Membuat matriks keputusan yang ternormalisasi terbobot. Dengan bobot $w_i = w_1, w_2, w_3, \dots, w_n$ dimana w_j adalah bobot dari kriteria ke-j dan $\sum_{j=1}^n w_j = 1$ maka normalisasi bobot matriks V seperti pada Persamaan 5.

$$v_{ij} = w_j \cdot r_{ij} \quad (2.5)$$

Dengan $i = 1, 2, 3, \dots, m$; dan $j = 1, 2, 3, \dots, n$; dimana v_{ij} adalah elemen dari matriks keputusan yang ternormalisasi terbobot V. w_j adalah bobot dari kriteria ke-j. r_{ij} adalah elemen dari matriks keputusan yang ternormalisasi R [4].

4. Menentukan matriks solusi ideal positif dan solusi ideal negatif Solusi ideal positif dinotasikan A^+ , sedangkan solusi ideal negatif dinotasikan A^- . Berikut ini adalah Persamaan 6 & 7 dari A^+ dan A^- :

$$A^+ = (y_1^+, y_2^+, \dots, y_n^+) \quad (2.6)$$

$$A^- = (y_1^-, y_2^-, \dots, y_n^-) \quad (2.7)$$

5. Menghitung Separasi

S^+ adalah jarak alternatif dari solusi ideal positif didefinisikan pada Persamaan 8.

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2} \quad (2.8)$$

S^- adalah jarak alternatif dari solusi ideal negatif didefinisikan pada Persamaan 9.

$$S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2} \quad (2.9)$$

Dimana:

S_i^+ adalah jarak alternatif ke-I dari solusi ideal positif,

S_i^- adalah jarak alternatif ke-I dari solusi ideal negatif,

v_{ij} adalah elemen dari matriks keputusan yang ternormalisasi terbobot V,

v_j^+ adalah elemen matriks solusi ideal positif,

v_j^- adalah elemen matriks solusi ideal negatif.

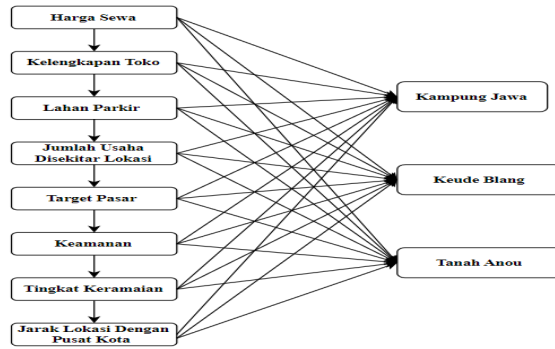
6. Menghitung kedekatan relatif terhadap solusi ideal positif c_i^+ pada persamaan 10 dibawah ini:

$$c_i^+ = \frac{S_i^-}{(S_i^- + S_i^+)} \quad (2.10)$$

7. Meranking Alternatif dengan melakukan pengurutan C^+ terbesar ke nilai terkecil. Alternatif dengan nilai C^+ terbesar merupakan solusi yang terbaik.

III. HASIL DAN PEMBAHASAN

3.1 Analisis Metode Hybrid DSS



Gambar 2. Struktur Hirarki AHP

Pada gambar diatas menjelaskan bahwa struktur hirarki menggambarkan pengolahan data dengan metode AHP dengan melihat relasi antara masalah kriteria dan alternatif. Masalah yang akan diatasi adalah melakukan pemilihan lokasi usaha dengan 8 kriteria yang terdapat pada Gambar 2. untuk Alternatif di inisialkan dengan A1, A2, dan A3. Setelah struktur hirarki terbentuk maka proses dilanjutkan untuk membuat matriks perbandingan berpasangan (*Pairwise Comparison*) Kriteria. Adapun matriks berpasangan yang terbentuk dapat dilihat pada Tabel 4 sebagai berikut:

Kriteria	Harga Sewa	Kelengkapan Toko	Lahan Parkir	Jumlah Usaha Disekitar Lokasi	Target Pasar	Keamanan	Tingkat Keramaian	Jarak Lokasi Dengan Pusat Kota
Harga Sewa	1	3	2	1	2	1	3	4
Kelengkapan Toko	0,333	1	2	2	2	1	2	2
Lahan Parkir	0,5	0,5	1	2	1	5	2	3
Jumlah Usaha Disekitar Lokasi	1	0,5	0,5	1	2	3	4	3
Target Pasar	0,5	0,5	1	0,5	1	2	3	2
Keamanan	1	1	0,2	0,333	0,5	1	1	3
Tingkat Keramaian	0,333	0,5	0,5	0,25	0,333	1	1	2
Jarak Lokasi Dengan Pusat Kota	0,25	0,5	0,333	0,333	0,5	0,333	0,5	1
Jumlah	4,916666667	7,5	7,533	7,41	9,3	14,3	16,5	20

Tabel 4. Perbandingan Berpasangan

Matriks perbandingan berpasangan yang terdapat pada Tabel 4.1, dimana masing – masing kriteria akan dibandingkan secara berpasangan berdasarkan tingkat kepentingannya dengan nilai yang diberikan menurut skala perbandingan 1 – 9.

Setelah matriks perbandingan berpasangan dihasilkan maka proses dilanjutkan kembali untuk menentukan nilai bobot antar kriteria dan prioritas. Dengan cara membagi isi matriks perbandingan dengan jumlah kolom yang beresuaian, kemudian menjumlahkan perbaris setelah itu hasil penjumlahan dibagi dengan banyaknya kriteria sehingga ditemukan bobot prioritas.

Kriteria	Harga Sewa	Kelengkapan Toko	Lahan Parkir	Jumlah Usaha Disekitar Lokasi	Target Pasar	Keamanan	Tingkat Keramaian	Jarak Lokasi Dengan Pusat Kota
Harga Sewa	0,203	0,4	0,256	0,135	0,210	0,07	0,182	0,2
Kelengkapan Toko	0,068	0,133	0,256	0,27	0,214	0,07	0,121	0,1
Lahan Parkir	0,102	0,677	0,133	0,27	0,107	0,349	0,121	0,15
Jumlah Usaha Disekitar Lokasi	0,203	0,677	0,066	0,135	0,214	0,209	0,242	0,15
Target Pasar	0,102	0,677	0,133	0,067	0,107	0,14	0,182	0,1
Keamanan	0,203	0,133	0,027	0,045	0,054	0,07	0,061	0,15
Tingkat Keramaian	0,068	0,677	0,066	0,034	0,036	0,07	0,061	0,1
Jarak Lokasi Dengan Pusat Kota	0,051	0,677	0,044	0,045	0,054	0,023	0,03	0,05

Tabel 5. Nilai Bobot Antar Kriteria Dan Prioritas

Tabel 5 merupakan tahapan dalam penentuan nilai prioritas bobot setiap kriteria, total semua nilai bobot adalah 1 atau 100. Setelah nilai bobot antar kriteria didapat maka proses masih dilanjutkan untuk

menghitung nilai vektor eigen dari setiap nilai kriteria perbandingan berpasangan. Adapun hasil tersebut dapat dilihat pada Tabel 6.

		Matriks Bobot Prioritas Kriteria						Jumlah
0,203309031	0,4	0,245486726	0,134831461	0,214289714	0,069767442	0,181818182	0,2	1,669579355
0,06779661	0,133333333	0,245486726	0,269662921	0,214289714	0,069767442	0,121212121	0,1	1,241544868
0,10494915	0,066666667	0,133743363	0,269662921	0,107142857	0,348837209	0,121212121	0,15	1,297960054
0,203309031	0,066666667	0,066371681	0,134831461	0,214289714	0,298932326	0,242424242	0,15	1,287271922
0,10494915	0,066666667	0,133743363	0,06741573	0,107142857	0,139534884	0,181818182	0,1	0,897016598
0,203309031	0,133333333	0,245486726	0,04494382	0,053571429	0,069767442	0,066666661	0,15	0,742160588
0,06779661	0,066666667	0,066371681	0,033797865	0,035714286	0,069767442	0,066666661	0,1	0,500630612
0,059847459	0,066666667	0,044247708	0,04494382	0,053571429	0,023255814	0,03030303	0,05	0,363836005

Tabel 6. Nilai Vektor Eigen

Untuk mendapatkan nilai vektor eigen berdasarkan proses perkalian antara nilai prioritas dan nilai matriks perbandingan berpasangan. Setelah matriks vektor eigen didapatkan maka proses perhitungan rasio dapat dilihat pada Tabel 7.

Jumlah	Prioritas	Hasil
1,669579355	0,208697419	8,847040174
1,241544868	0,155193108	9,048132353
1,297960054	0,162245007	9,267353604
1,287271922	0,160908999	8,810220255
0,897016598	0,112127075	8,927437646
0,742160588	0,092770073	8,600044913
0,500630612	0,062578826	8,824002233
0,363836005	0,045479501	8,822540081
		71,14677126

Tabel 7. Nilai Jumlah

Kolom jumlah dan kolom prioritas pada Tabel 7 diatas berasal dari langkah sebelumnya, pada langkah ini digunakan untuk mendapatkan nilai jumlah keseluruhan digunakan untuk langkah selanjutnya dibawah ini.

$$\lambda_{maks} = \frac{71,14677126}{8} = 8,893346407$$

$$CI = \frac{\lambda_{maks} - n}{n - 1} = \frac{8,893346407 - 8}{8 - 1} = \frac{0,893346407}{7} = 0,127620915$$

$$CR = \frac{0,127620915}{1,41} = 0,090511287$$

Oleh karena nilai CR < 0.1, maka nilai perbandingan kriteria berpasangan adalah konsisten dan dapat dilanjutkan ke langkah selanjutnya.

3.2 Proses Perhitungan Metode Topsis

Langkah 1: TOPSIS dimulai dengan membangun sebuah matriks keputusan. Matriks keputusan X mengacu terhadap m alternatif yang akan dievaluasi berdasarkan n kriteria. Matriks keputusan X dapat dilihat pada Tabel 8.

Alternatif	C1	C2	C3	C4	C5	C6	C7	C8
Kampung Jawa	5	3	4	4	4	3	4	3
Tanah Anou	5	3	3	3	2	4	2	4
Kende Blang	5	3	3	4	5	5	3	4

Tabel 8. Nilai Antar Kriteria Dan Prioritas

Tabel ini menampilkan hasil penilaian Pemeliti terhadap lokasi yang akan dijadikan tempat dalam pemilihan usaha.

Langkah 2: Membuat matriks keputusan ternormalisasi dari setiap kategori yang dilambangkan dengan X(i) dimana i=1, 2, 3, dan 4. Adapun rumus normalisasi matriks adalah sebagai berikut:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}$$

$$R = \frac{5}{\sqrt{5^2 + 5^2 + 5^2}} = \frac{5}{\sqrt{75}} = \frac{5}{8,660254038} = 0,577350269$$

	C1	C2	C3	C4	C5	C6	C7	C8
R =	0,57735	0,57735	0,5145	0,46852	0,29814	0,56569	0,37139	0,6247
	0,57735	0,57735	0,5145	0,6247	0,74536	0,70711	0,55709	0,6247

Hasil perhitungan dengan menggunakan rumus matriks keputusian ternormalisasi akan menghasilkan matriks normalisasi. Setelah didapatkan matriks keputusian maka proses perhitungan matriks keputusian ternormalisasi terbobot dapat dilihat pada perhitungan sebagai berikut:

W : [0,209; 0,155; 0,162; 0,161; 0,112; 0,093; 0,063; 0,045]

C1	C2	C3	C4	C5	C6	C7	C8
0,120491511	0,089600783	0,111299156	0,075389287	0,083574587	0,065598348	0,046482385	0,028410819
0,120491511	0,089600783	0,083474367	0,100519049	0,033429835	0,039359009	0,023241193	0,021308114
C1	C2	C3	C4	C5	C6	C7	C8
0,120491511	0,089600783	0,111299156	0,100519049	0,06685967	0,039359009	0,046482385	0,021308114
0,120491511	0,089600783	0,083474367	0,075389287	0,033429835	0,023478678	0,023241193	0,029410819
0,120491511	0,089600783	0,083474367	0,100519049	0,083574587	0,065598348	0,046482385	0,028410819

Pada langkah ini, nilai bobot dikalikan dengan nilai setiap kriteria, yang menghasilkan matriks V. Setelah matriks V didapat maka proses akan dilanjutkan untuk menentukan matriks solusi ideal positif (A+) dan solusi ideal negatif (A-) dengan melihat nilai tertinggi dan terendah yang dapat dilihat berikut ini:

C1	C2	C3	C4	C5	C6	C7	C8
0,120491511	0,089600783	0,111299156	0,075389287	0,083574587	0,065598348	0,046482385	0,028410819
0,120491511	0,089600783	0,083474367	0,100519049	0,033429835	0,039359009	0,023241193	0,021308114

Hasil yang disajikan pada Tabel 9 menggambarkan hasil nilai jarak solusi ideal positif dan negatif. Setelah hasil tersebut didapat maka tahapan dilanjutkan untuk menghitung keidekatan relative dari setiap alternatif terhadap solusi ideal positif (ci+). Hasil tersebut dapat dilihat pada Tabel 9.

	Positif	Negatif	Preferensi
A1	0,040618281	0,04931456	0,548348737
A2	0,06325341	0,029224632	0,316016984
A3	0,039252543	0,058210702	0,597257992

Tabel 9. Hasil Keidekatan Relatif Dan Alternatif Solusi

Tabel 9 menyajikan hasil keidekatan relative dari setiap alternatif. Tahap akhir proses ini melakukan proses perbandingan alternatif. Adapun hasil tersebut dapat dilihat pada Tabel 10.

Alternatif	Total	Rangking
Kampung Jawa	0,54835	2
Tanah Anou	0,31602	3
Keude Blang	0,59726	1

Tabel 10. Hasil Perbandingan Alternatif

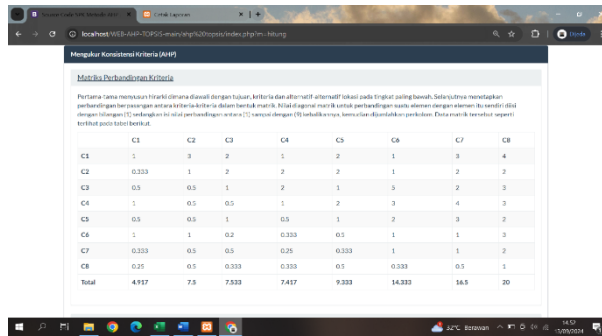
Hasil perbandingan atau rekomendasi lokasi usaha setelah melakukan Analisa Keude Blang menjadi lokasi usaha yang di pilih dan direkomendasikan dengan cara melakukan analisis dalam penentuan kriteria dan bobot yang telah dilakukan diatas.

3.3 Implementasi Hybrid DSS

Setelah mengolah data secara manual dengan menggunakan metode hybrid SPK, maka perlu dilakukan pengujian data dengan membangun aplikasi hybrid DSS menggunakan bahasa pemrograman PHP dan database MySQL. Dengan adanya aplikasi hybrid DSS ini dapat memudahkan dalam mengambil keputusan dengan cepat.

1. Halaman Menu Proses Metode AHP

Pada halaman tersebut terdapat nilai bobot yang di inputkan oleh admin. Nilai bobot tersebut diberikan berdasarkan skala perbandingan 1 – 9 yang telah ditetapkan. Untuk nilai kriteria yang sama tetap bernilai 1. Seperti pada gambar berikut :

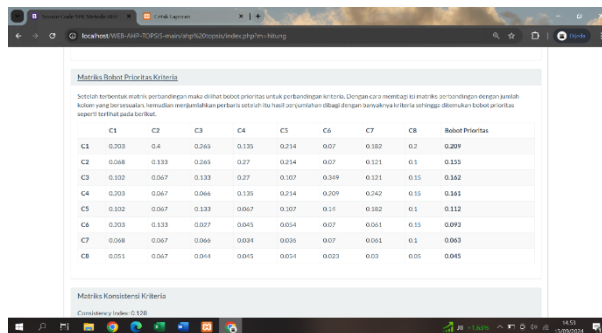


Gambar 3. Halaman Prosesi Metode AHP

Pada halaman tersebut dilakukan perhitungan nilai bobot kriteria menggunakan metode AHP dimana dilakukan perhitungan untuk mencari nilai eigen dan vector.

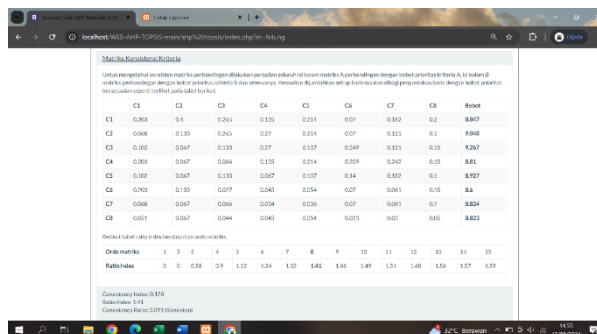
2. Halaman Matriks Bobot Prioritas Kriteria

Setelah terbentuk matriks perbandingan, maka dilihat bobot prioritas untuk perbandingan kriteria. Dengan cara membagi isi matriks perbandingan dengan jumlah kolom yang beresesuaian, kemudian menjumlahkan perbaris setelah itu hasil penjumlahan dibagi dengan banyaknya kriteria sehingga ditemukan bobot prioritas.



Gambar 4. Halaman Matriks Bobot Prioritas Kriteria

3. Halaman Matriks Konsistem Kriteria



Gambar 5. Halaman Matriks Konsisten Kriteria

Nilai pada matriks diatas konsisten karna nilai dari $CR < 0,1$ dan dapat dilanjutkan ke perhitungan selanjutnya yaitu melakukan analisa menggunakan metode topsis untuk menghasilkan solusi alternatif dengan melakukan perbandingan.

4. Perhitungan Topsis

	Harga Sewa	Kelengkapan Toko	Lahan Parkir	Jumlah Usaha Disekitar Lokasi	Target Pasar	Keamanan	Tingkat Keramaian	Jarak Lokasi Dengan Pusat Kota
Kampung Jawa	5	3	4	4	4	3	4	3
Tanah Anou	5	3	3	3	2	4	2	4
Keude Blang	5	3	3	4	5	5	3	4

Gambar 6. Perhitungan Topsis

Pada gambar tersebut menampilkan nilai bobot pada setiap alternatif berdasarkan observasi dan wawancara.

4. Halaman Normalisasi Matriks Keputusan

	C1	C2	C3	C4	C5	C6	C7	C8
A1	0.57735	0.57735	0.68599	0.6247	0.59628	0.42426	0.74278	0.46852
A2	0.57735	0.57735	0.5145	0.46852	0.29814	0.56569	0.37139	0.6247
A3	0.57735	0.57735	0.5145	0.6247	0.74536	0.70711	0.55709	0.6247

Gambar 7. Halaman Normalisasi Matriks Keputusan

Pada halaman tersebut dilakukan matriks perbandingan normalisasi dengan membagi dan dan memangkat nilai bobot pada setiap kolom dan barisnya.

5. Halaman Normalisasi Terbobot

	C1	C2	C3	C4	C5	C6	C7	C8
A1	0.12049	0.0896	0.1113	0.10052	0.06686	0.03936	0.04648	0.02131
A2	0.12049	0.0896	0.08347	0.07539	0.03343	0.05248	0.02234	0.02841
A3	0.12049	0.0896	0.08347	0.10052	0.08357	0.0656	0.03486	0.02841

Gambar 8. Halaman Normalisasi Terbobot

Halaman diatas menampilkan nilai yang sudah di normalisasi dengan cara membagi setiap kolom dengan bobot prioritas kriteria yang didapat menggunakan analisis metode AHP sebelumnya.

6. Halaman Matriks Solusi Ideal

	C1	C2	C3	C4	C5	C6	C7	C8
positif	0.12049	0.0896	0.1113	0.07539	0.08357	0.0656	0.04648	0.02841
negatif	0.12049	0.0896	0.08347	0.10052	0.03343	0.03936	0.02324	0.02131

	Positif	Negatif	Preferensi
A1	0.04062	0.04931	0.54835

Gambar 9. Halaman Matriks Solusi Ideal

Pada halaman tersebut yaitu menentukan jarak solusi ideal positif dan negatif. Berdasarkan nilai tertinggi dan terendah sesuai dengan cost dan benefit kriteria yang telah ditetapkan sebelumnya.

7. Halaman Jarak Solusi dan Nilai Preferensi

	Positif	Negatif	Preferensi
A1	0.04062	0.04931	0.54835
A2	0.06325	0.02922	0.31602
A3	0.03925	0.05821	0.59726

	Total	Rank
A1 - Kampung Jawa	0.548	2
A2 - Tanah Anou	0.316	3
A3 - Keude Blang	0.597	1

Gambar 10. Halaman Jarak Solusi dan Nilai Preferensi

Pada halaman tersebut menampilkan nilai solusi ideal positif dan negatif dengan cara memangkatkan dan mengurangi setiap baris dan kolomnya sesuai dengan yang telah ditentukan diatas

8. Halaman Perangkingan

	Total	Rank
A1 - Kampung Jawa	0.548	2
A2 - Tanah Anou	0.316	3
A3 - Keude Blang	0.597	1

Gambar 11. Halaman Perangkingan

Pada halaman di atas menampilkan hasil perankingan atau rekomendasi lokasi usaha setelah melakukan analisa. Kemudian Blang menjadi lokasi usaha yang di pilih dan direkomendasikan dengan cara melakukan analisis dalam penentuan kriteria dan bobot yang telah dilakukan diatas.

IV. KESIMPULAN

Analisis Hybrid Decision Support System (DSS) yang menggunakan metode AHP (Analytic Hierarchy Process) dan TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) merupakan pendekatan yang efektif dalam pengambilan keputusan pemilihan lokasi usaha.

- 1) AHP (Analytic Hierarchy Process): Metode ini digunakan untuk menentukan bobot atau tingkat kepentingan dari setiap kriteria berdasarkan preferensi atau kebijakan yang ditentukan oleh pengambil keputusan. Dalam konteks pemilihan lokasi usaha, kriteria seperti harga sewa, kelengkapan toko, lahan parkir, jumlah usaha di sekitar lokasi, target usaha, keamanan, tingkat keramaian, dan jarak lokasi dengan pusat kota dinilai untuk menentukan prioritas masing-masing.
- 2) TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution): Metode ini digunakan untuk menentukan peringkat alternatif (lokasi usaha) berdasarkan jaraknya dengan solusi ideal. TOPSIS membantu dalam mengidentifikasi lokasi yang paling mendekati kondisi ideal sesuai dengan bobot yang telah ditentukan oleh AHP.

Dengan menggabungkan AHP dan TOPSIS, analisis hybrid ini mampu memberikan rekomendasi lokasi usaha yang optimal berdasarkan nilai perankingan terhadap berbagai kriteria yang relevan.

DAFTAR PUSTAKA

- [1] I. Naufal and N. Nurdin, "Sistem Pendukung Keputusan Penentuan Penyakit Pada Tanaman Terong Menggunakan Metode Simple Additive Weighting," *TECHSI-Jurnal Tek. Inform.*, vol. 12, no. 1, pp. 123–139, 2020.
- [2] F. Febrianti and S. Bemil, "Strategi Mempertahankan Loyalitas Pelanggan Pada Usaha Kuliner Di Kecamatan Beikayang," *Inov. Pembang. J. Kelitbangan*, vol. 11, no. 02, pp. 189–210, 2023.
- [3] N. A. Kurniasih and S. P. Astuti, "Implementasi Metode Analytical Hierarchy Process (AHP) Sebagai Solusi Alternatif Dalam Pemilihan Supplier Bahan Baku Kedelai (Studi Kasus Pabrik Tahu Sehat Sari)." UIN Surakarta, 2021.
- [4] D. Guiswandi, M. Yanto, M. Hafizh, and L. Mayola, "Analisis hybrid decision support system dalam penentuan status kelulusan mahasiswa," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1127–1136, 2021.
- [5] A. Asrianda, H. A. K. Aidilof, L. Rosnita, and Z. Zulfadli, "Review AHP dalam Fenomena Gelombang Ekonomi," *TECHSI - J. Tek. Inform.*, vol. 14, no. 1, p. 59, 2023, doi: 10.29103/teichsi.v14i1.12588.
- [6] I. Nisaa and A. Wibowo, "Penentuan Dosen Terbaik Menggunakan Metode Analytical Hierarchy Process (AHP) dan Technique For Order By Similarity To Ideal Solution (TOPSIS): Studi Kasus Akademi Teknologi Bogor," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 12, no. 2, pp. 62–74, 2020.

SENASTIKA Universitas Malikussaleh

ANALISIS PERBANDINGAN KINERJA ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK KLASTERISASI PRODUKSI PADI DI PULAU SUMATERA

Siti Nadilla^{*1}, Wirda Syahrifa², Fakhru Razi³, Rizki Maulana⁴, Munirul Ula⁵

^{1,2,3,4} Program Studi Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹siti.210170032@mhs.unimal.ac.id, ²wirda.210170033@mhs.unimal.ac.id,
³fakhru.210170029@mhs.unimal.ac.id, ⁴rizki.210170037@mhs.unimal.ac.id,
⁵munirulula@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan analisis pengelompokan data tanaman padi di Pulau Sumatra selama periode 2010 hingga 2020 dengan menerapkan algoritma K-Means dan K-Medoids. Pengelompokan ini penting untuk mengidentifikasi pola produksi padi, sehingga dapat membedakan daerah dengan hasil produksi tertinggi dan terendah. Data yang dianalisis mencakup variabel kunci seperti luas lahan pertanian, tingkat produktivitas, dan total hasil panen, yang diperoleh dari sumber resmi pertanian. Metodologi yang digunakan mencakup langkah-langkah preprocessing data, seleksi fitur yang relevan, dan penerapan algoritma pengelompokan. Hasil analisis menunjukkan bahwa algoritma K-Means menawarkan kecepatan dan efisiensi yang lebih tinggi dalam membentuk kluster berdasarkan karakteristik produksi, sementara K-Medoids menunjukkan ketahanan yang lebih baik terhadap data yang tidak biasa. Temuan dari pengelompokan ini berhasil mengidentifikasi sejumlah daerah dengan produksi padi tertinggi dan terendah, memberikan wawasan berharga bagi kepentingan pembaca dalam meningkatkan strategi pengelolaan sumber daya pertanian. Penelitian ini diharapkan dapat menjadi dasar bagi pengambilan keputusan yang lebih baik dalam upaya meningkatkan produktivitas dan keberlanjutan sektor pertanian di Pulau Sumatra.

Keywords: *Pengelompokan, K-Means, K-Medoids, Tanaman Padi, Sumatera.*

1. PENDAHULUAN

Pertanian padi di Indonesia, khususnya di Pulau Sumatra, merupakan salah satu sektor penting dalam mendukung ketahanan pangan nasional. Pulau Sumatra memiliki potensi lahan yang luas dan beragam, yang jika dikelola dengan baik dapat meningkatkan hasil produksi padi secara signifikan. Namun, meskipun terdapat berbagai sumber daya alam yang melimpah, disparitas dalam hasil produksi padi masih terjadi di berbagai daerah. Oleh karena itu, analisis pengelompokan data produksi padi menjadi penting untuk memahami faktor-faktor yang memengaruhi kinerja produksi dan untuk merumuskan strategi peningkatan produktivitas [1].

Padi (*Oryza sativa*) merupakan komoditas pangan utama di Indonesia, khususnya di Pulau Sumatra yang memiliki potensi besar dalam produksi beras. Sebagai salah satu sumber pangan utama, pengelolaan data terkait tanaman padi menjadi penting untuk mendukung peningkatan produktivitas dan efisiensi pertanian [1]. Salah satu pendekatan yang dapat digunakan untuk analisis data tanaman padi adalah pengelompokan data atau *clustering*, yang bertujuan untuk mengidentifikasi pola dan mengelompokkan area berdasarkan karakteristik yang relevan seperti jenis lahan, cuaca, dan hasil panen [2].

Algoritma K-Means dan K-Medoids merupakan dua metode pengelompokan yang sering digunakan dalam analisis data. Algoritma K-Means bekerja dengan meminimalkan variasi dalam kluster dengan menghitung centroid, yaitu titik tengah dari setiap kluster yang terbentuk [3]. Sementara itu, K-Medoids bekerja dengan memilih titik data yang aktual sebagai pusat kluster, sehingga lebih robust terhadap outlier dibandingkan K-Means [4]. Penggunaan kedua algoritma ini sangat relevan dalam konteks pengelolaan data tanaman padi, karena seringkali data pertanian memiliki distribusi yang tidak merata dan rentan terhadap outlier [5].

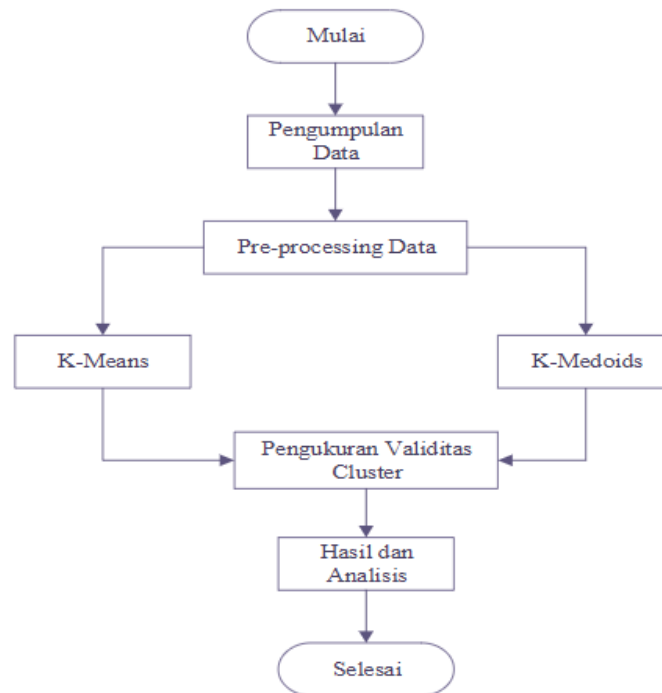
Pulau Sumatra memiliki variasi kondisi agroklimat yang mempengaruhi hasil panen padi di setiap wilayahnya. Dengan demikian, analisis clustering dapat membantu mengelompokkan wilayah-wilayah yang memiliki kesamaan karakteristik dan mengidentifikasi faktor-faktor yang mempengaruhi produktivitas padi. Studi sebelumnya telah menunjukkan bahwa teknik pengelompokan berbasis algoritma seperti K-Means dan K-Medoids efektif dalam mengidentifikasi pola tersembunyi dalam data pertanian, serta memberikan informasi yang bermanfaat untuk pengambilan keputusan di sektor pertanian.

Penelitian ini bertujuan untuk melakukan analisis pengelompokan data tanaman padi di Pulau Sumatra menggunakan algoritma K-Means dan K-Medoids. Dengan mengimplementasikan dua algoritma tersebut dengan mencoba beberapa kluster untuk mengelompokkan wilayah hasil produksi tanaman padi di Provinsi Sumatera, yang lebih jelas mengenai distribusi dan potensi wilayah tanam padi di Sumatera [6]. Pada hasil akhir penelitian, kedua algoritma, K-Means dan K-Medoids, berhasil mengelompokkan data dengan baik. Namun, hasil

menunjukkan bahwa algoritma K-Medoids memberikan performa yang lebih baik dibandingkan K-Means. Algoritma K-Medoids mampu menghasilkan kelompok yang lebih stabil dan akurat. Selain itu, hasil dari studi ini juga diharapkan dapat memberikan wawasan bagi pengambil kebijakan dalam merancang strategi peningkatan produktivitas pertanian yang lebih efektif dan efisien [7].

2. METODE PENELITIAN

Penelitian ini akan melalui lima tahap utama, yaitu pengumpulan data, kemudian dataset yang diperoleh akan melalui proses *pre-processing*. Data yang telah melalui tahap peningkatan kualitas akan diproses menggunakan dua algoritma, yaitu K-Means dan K-Medoids. Setelah data diproses oleh masing-masing algoritma, validitas kluster akan dihitung menggunakan metode DBI. Hasil klustering dan perhitungan validitas DBI kemudian akan dianalisis untuk mendapatkan kesimpulan. Metodologi penelitian ini dapat dilihat pada gambar 1.

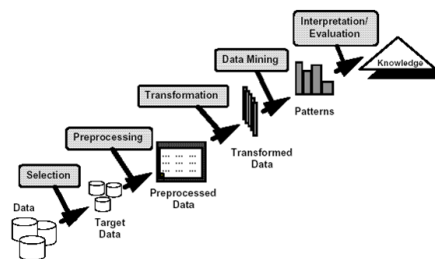


Gambar 1. Metodologi Penelitian

Dalam penelitian ini, dataset yang digunakan adalah data tanaman padi di pulau Sumatera yang diperoleh dari kaggle.com, dengan rentang waktu dari tahun 2010 hingga tahun 2020. Data tersebut akan melalui tahap *pre-processing*, dimulai dengan pembersihan data (cleaning). Setelah kualitas data dianggap memadai, data akan diproses menggunakan dua algoritma, yaitu K-Means dan K-Medoids. Tahap berikutnya adalah analisis hasil, di mana validitas kluster dihitung menggunakan Davies Bouldin Index (DBI) untuk menentukan algoritma terbaik.

2.1 Data Mining

Data mining adalah proses yang digunakan untuk menemukan pola, hubungan, atau informasi penting dari sejumlah besar data yang tersebar secara tidak terstruktur. Melalui data mining, teknik-teknik analisis digunakan untuk mengekstrak informasi tersembunyi yang dapat membantu pengambilan keputusan dan pengembangan model prediksi [8]. Menurut Han et al. (2019), data mining melibatkan serangkaian proses seperti pengumpulan data, pembersihan data, transformasi, dan penambangan itu sendiri, yang bertujuan untuk menghasilkan pola yang bermakna dari kumpulan data yang besar. Teknik-teknik yang digunakan dalam data mining meliputi algoritma klasifikasi, clustering, regresi, asosiasi, dan analisis deret waktu.



Gambar 2. Proses Data Mining

2.2 Clustering

Clustering adalah salah satu teknik dalam data mining yang memiliki tujuan untuk mengelompokkan objek-objek ke dalam kelompok atau *cluster* berdasarkan kesamaan karakteristiknya. Objek-objek yang berada dalam satu kelompok memiliki karakteristik yang lebih mirip satu sama lain dibandingkan dengan objek-objek yang berada di kelompok lain. Proses clustering digunakan untuk menemukan struktur tersembunyi dalam data yang tidak memiliki label atau kategori sebelumnya, sehingga dapat memberikan pemahaman lebih baik tentang data tersebut.

2.3 K-Means

K-Means adalah salah satu algoritma *clustering* yang paling populer dan sederhana dalam data mining. Algoritma ini bertujuan untuk membagi sekumpulan data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik. Setiap *cluster* direpresentasikan oleh sebuah titik pusat, yang disebut centroid, yang merupakan rata-rata dari semua data dalam *cluster* tersebut. Algoritma K-Means bekerja dengan cara meminimalkan jarak antara data dengan centroid *cluster*-nya [9]. Dalam kata lain, K-Means mengelompokkan data sedemikian rupa sehingga objek-objek yang mirip berada dalam kelompok yang sama, dan objek-objek yang berbeda ditempatkan dalam kelompok yang berbeda. Ada beberapa langkah dalam melakukan perhitungan menggunakan algoritma K-Means yaitu:

1. Tentukan nilai K untuk titik awal klaster secara acak
2. Setiap data akan dibagi ke dalam klaster berdasarkan jarak Euclidean ke titik pusat klaster. Rumus untuk menghitung jarak Euclidean dapat dilihat pada persamaan. (1):

$$d(x,y) = \|x-y\|^2 = \dots\dots\dots(1)$$

dengan :

d= jarak data ke pusat cluster
 x= data ke pada atribut ke
 y= titik pusat ke, pada atribut

3. Pusat klaster akan dihitung kembali berdasarkan nilai rata-rata dari data yang ada dalam klaster tersebut.
4. Proses di tahap 2 dan 3 akan diulang hingga ditemukan kelompok yang stabil, di mana tidak ada perubahan anggota dalam klaster dibandingkan dengan percobaan sebelumnya.

2.4 K-Medoids

K-Medoids adalah algoritma *clustering* yang mirip dengan K-Means, tetapi memiliki perbedaan mendasar dalam cara menentukan pusat klaster. Sementara K-Means menggunakan centroid (yang merupakan rata-rata dari semua titik dalam kluster) sebagai representasi dari setiap kluster, K-Medoids menggunakan titik data aktual yang terletak di dalam kluster tersebut sebagai pusat kluster, yang dikenal sebagai medoid. Dengan demikian, K-Medoids lebih robust terhadap outlier dan noise dalam data, karena medoid adalah titik yang paling representatif dari sekumpulan data dalam kluster [10].

Tahap-tahap penyelesaian K-Medoids adalah sebagai berikut:

1. Tentukan jumlah K sebagai pusat klaster.
2. Setiap data dimasukkan ke dalam klaster terdekat dengan cara mengukur jarak menggunakan rumus Jarak Euclidean sesuai dengan persamaan (1).
3. Pilih secara acak satu data dari setiap klaster untuk dijadikan medoid yang baru.
4. Objek yang termasuk dalam masing-masing klaster akan dihitung menggunakan medoid yang baru.
5. Total simpangan akan dihitung dengan membandingkan total jarak baru dengan total jarak awal. Jika hasilnya $S < 0$, cari data baru untuk membentuk k objek yang baru sebagai medoid.
6. Ulangi langkah 3 hingga medoid tidak lagi berubah, sehingga diperoleh anggota untuk masing-masing klaster.[11]

2.5 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) adalah metrik yang digunakan untuk mengevaluasi kualitas clustering yang dihasilkan oleh algoritma pengelompokan. DBI mengukur sejauh mana setiap kluster terpisah dari kluster lainnya, serta seberapa baik kluster tersebut terdefinisi. Nilai DBI yang lebih rendah menunjukkan bahwa kluster yang terbentuk memiliki jarak yang lebih besar satu sama lain (artinya, mereka lebih terpisah) dan bahwa titik-titik dalam kluster tersebut lebih dekat satu sama lain (artinya, kluster tersebut lebih padat).

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data dan Pre-Processing Data

Pada penelitian ini, Proses pengumpulan data merupakan tahap awal yang sangat penting dalam analisis data. Data yang digunakan dalam penelitian ini diperoleh kaggle terkait produksi padi di Pulau Sumatra selama periode 2010 hingga 2020. Sumber data mencakup laporan tahunan produksi padi, serta luas panen.

Tabel 1. Data awal produksi padi

No	Provinsi	Produksi	Luas Panen
1.	Aceh 2010	1788738	352281
2.	Aceh 2011	1772962	380686
3.	Aceh 2012	1582393	387803
4.	Aceh 2013	2331046	419183
5.	Aceh 2014	1820062	376137
6.	Aceh 2015	1956940	461060
7.	Aceh 2016	2180754	293067
8.	Aceh 2017	2478922	294483
9.	Aceh 2018	1751996,94	329515,78
10.	Aceh 2019	1714437,6	310012,46
11.	Aceh 2020	1861567,1	317869,41
12.	Sumatera Utara 2010	3582302	754674
13.	Sumatera Utara 2011	3607403	757547
14.	Sumatera Utara 2012	3715514	765099
15.	Sumatera Utara 2013	3727249	742968
...
87.	Lampung 2019	2164089,33	464103,42
88.	Lampung 2020	2604913,29	545149,05

Data yang digunakan dalam penelitian ini telah melalui proses normalisasi menggunakan teknik Min-Max Normalization. Teknik ini mengubah nilai-nilai asli setiap variabel ke dalam rentang 0 hingga 1, sehingga memungkinkan setiap fitur memiliki skala yang sebanding dan mencegah bias pada hasil pengelompokan

Tabel 2. Hasil Data Normalisasi

No	Provinsi	Produksi	Luas Panen
1.	Aceh 2010	0,335	0,357
2.	Aceh 2011	0,332	0,392
3.	Aceh 2012	0,291	0,401
4.	Aceh 2013	0,452	0,440
5.	Aceh 2014	0,342	0,387
6.	Aceh 2015	0,371	0,492
7.	Aceh 2016	0,419	0,284
8.	Aceh 2017	0,483	0,286
9.	Aceh 2018	0,327	0,329
10.	Aceh 2019	0,3192	0,305
11.	Aceh 2020	0,351	0,315
12.	Sumatera Utara 2010	0,721	0,854
13.	Sumatera Utara 2011	0,726	0,858
14.	Sumatera Utara 2012	0,749	0,867
15.	Sumatera Utara 2013	0,752	0,840
...
87.	Lampung 2019	0,416	0,495
88.	Lampung 2020	0,511	0,595

3.2 Proses Klasterisasi dan Validitas

Proses klasterisasi atau pengelompokan pada penelitian ini akan dilakukan dengan menggunakan dua algoritma, yaitu algoritma K-Means dan algoritma K-Medoids. Model ini memproses 88 data tanaman padi di Pulau Sumatera dari tahun 2010 sampai 2020 yang berisi hasil panen dan luas lahan.

K-Means

Proses klasterisasi pada penelitian ini menggunakan model K-Means. Percobaan klaster dilakukan dengan menggunakan jumlah klaster 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Dari hasil analisis, klaster ke-3 memberikan hasil terbaik dalam memisahkan data secara optimal. Hasil dari klasterisasi tersebut dapat dilihat pada tabel dibawah ini.

Tabel 3. Hasil Pengelompokan klaster pertama K-Means

No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Sumatera Utara 2010	0.721	0.854	1
2.	Sumatera Utara 2011	0.726	0.858	1
3.	Sumatera Utara 2012	0.749	0.867	1
4.	Sumatera Utara 2013	0.751	0.840	1
5.	Sumatera Utara 2014	0.731	0.808	1
6.	Sumatera Utara 2015	0.820	0.888	1
7.	Sumatera Utara 2016	0.894	0.445	1
...
20.	Lampung 2012	0.617	0.715	1
21.	Lampung 2013	0.640	0.710	1
22.	Lampung 2014	0.664	0.723	1
23.	Lampung 2015	0.734	0.796	1
24.	Lampung 2016	0.774	0.405	1
25.	Lampung 2017	0.830	0.412	1

Tabel 4. Hasil Pengelompokan klaster kedua K-Means

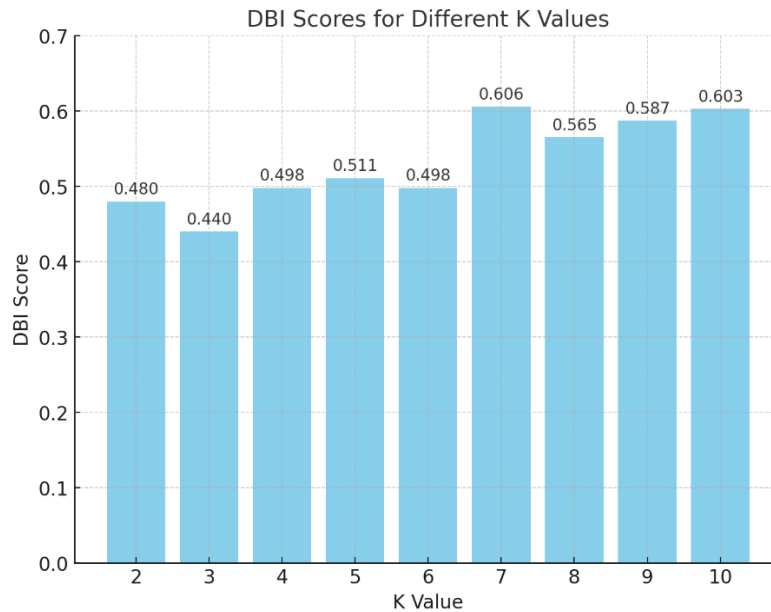
No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Aceh 2010	0.335	0.357	2
2.	Aceh 2011	0.332	0.392	2
3.	Aceh 2012	0.291	0.401	2
4.	Aceh 2013	0.452	0.440	2
5.	Aceh 2014	0.342	0.387	2
6.	Aceh 2015	0.371	0.492	2
7.	Aceh 2016	0.419	0.284	2
...
25.	Sumatera Barat 2020	0.262	0.287	2
26.	Sumatera Selatan 2019	0.510	0.588	2
27.	Sumatera Selatan 2020	0.530	0.603	2
28.	Lampung 2018	0.486	0.554	2
29.	Lampung 2019	0.416	0.495	2
30.	Lampung 2020	0.511	0.595	2

Tabel 5. Hasil Pengelompokan klaster ketiga K-Means

No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Riau 2010	0.074	0.115	3
2.	Riau 2011	0.066	0.101	3
3.	Riau 2012	0.060	0.100	3
4.	Riau 2013	0.044	0.068	3
5.	Riau 2014	0.0333	0.053	3
6.	Riau 2015	0.035	0.055	3
7.	Riau 2016	0.020	0.011	3
...
27.	Bengkulu 2014	0.078	0.104	3
28.	Bengkulu 2015	0.075	0.081	3
29.	Bengkulu 2016	0.086	0.025	3
30.	Bengkulu 2017	0.104	0.024	3
31.	Bengkulu 2018	0.012	0.003	3

32.	Bengkulu 2019	0.014	0.002	3
33.	Bengkulu 2020	0.014	0.001	3

Selanjutnya, dilakukan analisis terhadap 3 cluster tersebut untuk menentukan label wilayah produksi tinggi, sedang, dan rendah. Penentuan label ini didasarkan pada karakteristik data provinsi di setiap cluster. Setelah melakukan proses klasterisasi menggunakan model K-Means, validasi klaster akan dilakukan dengan menggunakan teknik DBI. Nilai DBI yang diperoleh berdasarkan hasil klasterisasi data menggunakan algoritma K-Means ditunjukkan pada Gambar 3.



Gambar 3. Nilai DBI K-Means

Berdasar pada nilai DBI terbaik, algoritma K-Means menunjukkan bahwa klaster yang optimal terjadi pada pengujian kalster 3, dengan nilai DBI sebesar 0,440. Pengujian ini membagi data menjadi 3 cluster, di mana cluster 1 memiliki jumlah sebanyak 25 wilayah untuk hasil produksi tinggi, cluster 2 memiliki total 30 wilayah untuk hasil produksi sedang, dan cluster 3 terdiri dari 33 wilayah untuk hasil produksi rendah. elemen,

K-Medoids

Proses pengelompokan dalam penelitian ini juga menggunakan algoritma K-Medoid, yang menerapkan uji kelompok yang sama seperti pada algoritma sebelumnya. Hasil dari klasterisasi dapat dilihat pada tabel dibawah.

Tabel 6. Hasil Pengelompokkan klaster pertama K-Medoids

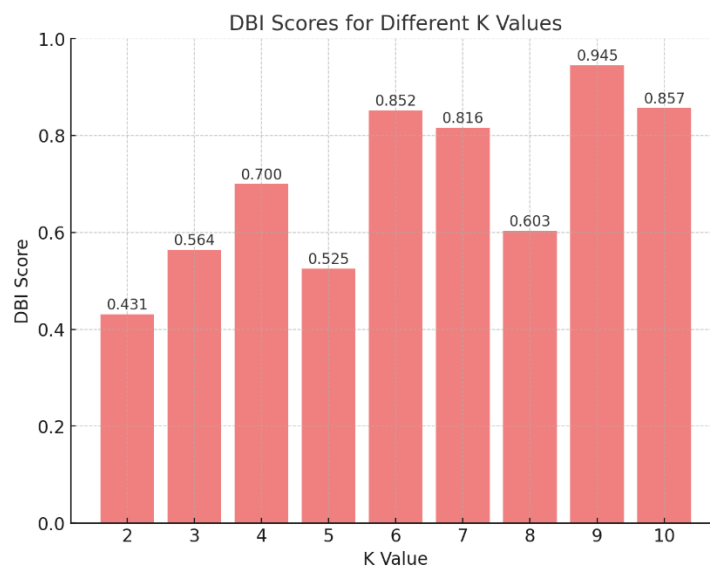
No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Sumatera Selatan 2014	0.740	0.924	1
2.	Sumatera Selatan 2015	0.864	1	1
3.	Sumatera Selatan 2016	1	0.682	1
4.	Sumatera Selatan 2017	0.984	0.690	1
5.	Sumatera Selatan 2018	0.594	0.640	1
6.	Sumatera Selatan 2019	0.510	0.588	1
7.	Sumatera Selatan 2020	0.530	0.603	1
...
50.	Sumatera Utara 2013	0.752	0.840	1
51.	Sumatera Utara 2014	0.731	0.808	1
52.	Sumatera Utara 2015	0.820	0.888	1
53.	Sumatera Utara 2016	0.894	0.445	1
54.	Sumatera Utara 2017	0.955	0.435	1

Tabel 7. Hasil Pengelompokkan klaster kedua K-Medoids

No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Sumatera Barat 2020	0.262	0.287	2
2.	Riau 2010	0.074	0.115	2

3.	Riau 2011	0.066	0.101	2
4.	Riau 2012	0.060	0.100	2
5.	Riau 2013	0.044	0.068	2
6.	Riau 2014	0.033	0.053	2
7.	Riau 2015	0.035	0.055	2
...
29.	Bengkulu 2015	0.075	0.081	2
30.	Bengkulu 2016	0.086	0.025	2
31.	Bengkulu 2017	0.104	0.024	2
32.	Bengkulu 2018	0.012	0.003	2
33.	Bengkulu 2019	0.014	0.002	2
34.	Bengkulu 2020	0.014	0.001	2

Selain itu, validasi kluster juga dilakukan dengan menggunakan Davies Bouldin Index (DBI). Nilai DBI yang didapat dari hasil klusterisasi data dengan algoritma K-Medoid dilihat pada Gambar 4.

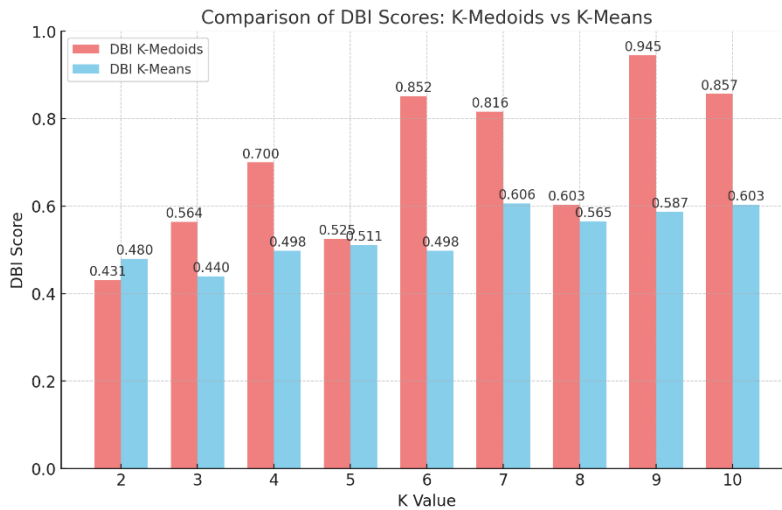


Gambar 4. Nilai DBI K-Medoids

Berdasarkan nilai DBI terbaik, algoritma K-Medoids menunjukkan bahwa cluster yang optimal terjadi pada pengujian kluster 2, dengan nilai DBI sebesar 0,431. Dengan kluster ke-2, data dibagi menjadi dua kelompok, di mana kelompok pertama berjumlah 54 wilayah dengan hasil produksi tinggi dan kelompok kedua terdiri dari 34 wilayah dengan hasil produksi rendah.

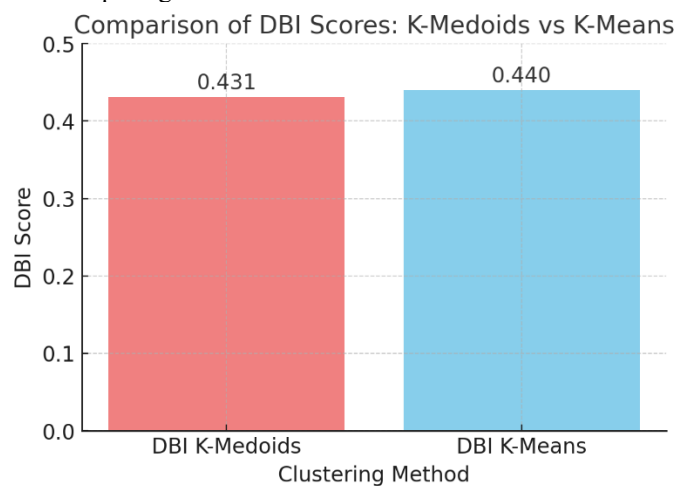
3.3 Perbandingan Kinerja Algoritma K-Means dan K-medoids

Perbandingan antara algoritma K-Means dan K-Medoids pada klusterisasi data tanaman padi di Pulau Sumatera dapat dilihat pada gambar 5.



Gambar 5. Perbandingan Nilai DBI K-Means dan K-Medoids

Berdasarkan hasil analisis dan pengolahan data, perbandingan antara algoritma K-Means dan K-Medoids menunjukkan perbedaan yang signifikan dalam kinerja keduanya. Algoritma K-Medoids berhasil menghasilkan cluster terbaik dengan nilai kluster ke-2, yang membagi data menjadi dua kelompok, yaitu kelompok 1 mencakup 54 wilayah dengan hasil produksi tinggi, sedangkan kelompok 2 terdiri dari 34 wilayah dengan hasil produksi rendah. Di sisi lain, algoritma K-Means menunjukkan hasil yang optimal pada K=3. Perbandingan DBI terbaik K-Means dan K-Medoids dapat dilihat pada gambar 6 dibawah ini:



Gambar 6. Perbandingan Nilai DBI Terbaik K-Means dan K-Medoids

Validasi klusterisasi dilakukan menggunakan Davies Bouldin Index (DBI), di mana K-Medoids menunjukkan nilai DBI terbaik sebesar 0,43, lebih rendah dibandingkan K-Means yang memiliki nilai DBI 0,440. Hal ini menegaskan bahwa K-Medoids lebih baik dalam mengelompokkan data dalam konteks penelitian ini, karena dapat menciptakan kluster yang lebih terpisah dan homogen.

4. DISKUSI

Berdasarkan hasil analisis klusterisasi pada data produksi tanaman padi di Pulau Sumatera dari tahun 2010 hingga 2020, penelitian ini membandingkan dua algoritma klusterisasi yang umum digunakan, yaitu K-Means dan K-Medoids. Proses klusterisasi dilakukan untuk mengelompokkan wilayah berdasarkan produksi padi mereka.

4.1 Perbandingan Dengan Penelitian Terdahulu

Hasil penelitian ini yang membahas klusterisasi data produksi padi di Pulau Sumatera menunjukkan bahwa algoritma K-Medoids dengan K=2 menghasilkan kluster terbaik, dengan nilai Davies Bouldin Index (DBI) sebesar 0,43. Di sisi lain, algoritma K-Means menghasilkan kluster terbaik dengan K=3, namun dengan nilai DBI yang sedikit lebih tinggi, yaitu 0,440. Temuan ini dapat dibandingkan dengan hasil penelitian yang serupa pada data persediaan di PT. Asia Petrocom Services, yang juga menggunakan algoritma K-Means dan K-Medoids untuk analisis klusterisasi.

Pada penelitian PT. Asia Petrocom Services, K-Medoids juga menghasilkan kluster terbaik dengan $K=2$, namun dengan nilai DBI yang jauh lebih rendah, yaitu 0,073. Sementara itu, K-Means menghasilkan kluster terbaik dengan $K=4$, menunjukkan bahwa pada data persediaan yang besar dan kompleks, K-Means cenderung memerlukan lebih banyak kluster untuk mencapai hasil yang optimal. Namun, walaupun jumlah kluster lebih banyak, K-Means tetap tidak mengungguli K-Medoids dalam hal kualitas klusterisasi, sebagaimana diukur dengan nilai DBI.

4.2 Keterbatasan Penelitian

Penelitian ini, meskipun berhasil memberikan hasil klusterisasi yang baik menggunakan algoritma K-Means dan K-Medoids, memiliki beberapa keterbatasan yang perlu diperhatikan seperti Data yang digunakan dalam penelitian ini terbatas pada periode 2010-2020 dan hanya mencakup wilayah di Pulau Sumatera. Dengan demikian, hasil penelitian ini tidak dapat digeneralisasi untuk seluruh wilayah di Indonesia atau untuk periode waktu yang lebih Panjang dan penelitian ini hanya berfokus pada data produksi padi tanpa mempertimbangkan faktor-faktor eksternal seperti cuaca, kebijakan pertanian, atau infrastruktur irigasi yang mungkin memiliki pengaruh signifikan terhadap produksi padi di setiap wilayah.

Meskipun penelitian ini memberikan hasil yang valid dan relevan, ada ruang untuk pengembangan lebih lanjut, terutama dalam hal variasi parameter, validasi yang lebih luas, serta penggunaan data dan algoritma yang lebih beragam. Hal ini akan memperkuat keakuratan hasil klusterisasi serta memperluas relevansi penelitian ini dalam konteks yang lebih luas.

5. KESIMPULAN

Berdasarkan hasil analisis dan pengolahan data tanaman padi di Pulau Sumatera dari tahun 2010 sampai tahun 2020 menggunakan algoritma K-Means dan K-Medoids, klasster terbaik diperoleh dengan algoritma K-Medoids pada nilai kluster ke-2. Sementara itu, algoritma K-Means menghasilkan kluster terbaik dengan kluster ke-3. Selain menghasilkan kluster yang lebih kompak, K-Medoids juga berhasil mengidentifikasi dua kelompok wilayah berdasarkan produksi padi. Hasil klusterisasi tersebut divalidasi menggunakan Davies Bouldin Index (DBI). Dalam penelitian ini, K-Medoids menunjukkan kinerja yang lebih baik dibandingkan K-Means, dengan nilai DBI terbaik sebesar 0,43 untuk K-Medoids dan 0,440 untuk K-Means. Pengujian K-Medoids dengan $K=2$ mengelompokkan data menjadi 2 kelompok, di mana kelompok 1 terdiri dari 54 wilayah dengan hasil produksi tinggi dan kelompok 2 terdiri dari 34 wilayah dengan hasil produksi rendah.

DAFTAR PUSTAKA

- [1] Iyan, R., (2020). ANALISIS KOMODITAS UNGGULAN SEKTOR PERTANIAN DI WILAYAH SUMATERA. *JURNAL SOSIAL EKONOMI PEMBANGUNAN*, 11(4), 215-235.
- [2] Yulianto, T., & Susilo, A. (2020). Penggunaan Algoritma K-Means untuk Segmentasi Data Pertanian di Sumatera. *Jurnal Teknik Pertanian Indonesia*, 11(2), 85-94.
- [3] Das, D., & Pradhan, S. (2020). Application of K-means and K-medoids Clustering Algorithms for Data Mining in Agriculture. *Journal of Data Mining and Knowledge Discovery*, 14(3), 214-228.
- [4] Wang, X., Chen, L., & Zhou, F. (2020). Comparative Study of K-Means and K-Medoids Clustering in Agricultural Data Analysis. *Computational Agriculture*, 19(2), 110-118.
- [5] Nurdin, I., & Fatimah, N. (2019). Implementing K-Means and K-Medoids for Agricultural Data Grouping. *Journal of Agricultural Data Science*, 6(4), 198-207.
- [6] Puspita, M., & Ningsih, S. (2021). Penggunaan Metode Clustering dalam Analisis Data Pertanian: Studi Kasus di Sumatera. *Jurnal Informatika dan Sains Data*, 15(4), 210-222.
- [7] Putra, I. R., & Wahyudi, A. (2020). Analisis Pengelompokan Data Produksi Padi di Sumatera Menggunakan Algoritma K-Medoids. *Jurnal Teknik Informatika dan Sistem Informasi*, 10(1), 45-55.
- [8] Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons.
- [9] Han, J., Pei, J., & Kamber, M. (2019). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- [10] Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2), 3336-3341.
- [11] Mubarak, H.A. Penerapan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Data Inventaris Rig. *IJRSE: Indonesian Journal of Informatic Research and Software Engineering*. 3(2), 127-135.

IMPLEMENTASI METODE *PROMETHEE* II DALAM PRODUKTIVITAS SAYURAN UNGGULAN PADA KECAMATAN DI ACEH UTARA

Al Kautsar¹, Nufus Fatayati², Rini Meiyanti³

¹²³Program Studi Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email : ¹kautsar.210170015@mhs.unimal.ac.id, ²nufus.210170023@mhs.unimal.ac.id,
³rinimeiyanti@unimal.ac.id

ABSTRAK

Penelitian ini bertujuan untuk menentukan kecamatan dengan produktivitas sayuran unggulan di Aceh Utara menggunakan Metode *PROMETHEE* II. Metode ini digunakan dalam sistem pendukung keputusan untuk membantu menganalisis berbagai kriteria yang terkait dengan produktivitas sayuran, seperti luas tanam, luas panen, luas tambah tanam, dan jumlah produksi. Data penelitian diambil dari 23 kecamatan di Aceh Utara yang mencakup berbagai jenis sayuran, seperti cabai, tomat, kacang panjang, dan terung, selama tahun 2023. Melalui analisis menggunakan Metode *PROMETHEE*, penelitian ini menghasilkan peringkat kecamatan berdasarkan nilai *Net Flow* yang mencerminkan tingkat produktivitas masing-masing kecamatan. Hasilnya menunjukkan bahwa Kecamatan Tanah Luas memiliki produktivitas sayuran dengan nilai *Net Flow* sebesar 3.0608, sedangkan Kecamatan Tanah Pasir berada pada posisi terendah dengan nilai -2.2172. Dengan hasil ini, penelitian memberikan rekomendasi kepada pemerintah daerah dan kelompok tani untuk fokus meningkatkan produktivitas di kecamatan dengan peringkat rendah. Kesimpulannya, Metode *PROMETHEE* II memberikan hasil yang akurat dan objektif dalam menganalisis produktivitas sayuran di berbagai kecamatan dan dapat dijadikan alat bantu dalam pengambilan keputusan terkait strategi peningkatan pertanian sayuran di Aceh Utara. Oleh sebab itu, hasil dari penelitian ini diharapkan bisa dijadikan pedoman dalam perencanaan strategi peningkatan produktivitas pertanian di Aceh Utara.

Kata Kunci : *PROMETHEE* II, Produktivitas Sayuran, Sistem Pendukung Keputusan, Kecamatan, Aceh Utara, *Net Flow*

ABSTRACT

This research aims to determine the districts with superior vegetable productivity in North Aceh using the PROMETHEE II method. This method is applied in a decision support system to assist in analyzing various criteria related to vegetable productivity, such as planting area, harvested area, additional planting area, and production volume. The research data were collected from 23 districts in North Aceh, covering various types of vegetables, such as chili, tomato, long bean, and eggplant, during 2023. Through analysis using the PROMETHEE method, this study produces a ranking of districts based on Net Flow values that reflect the productivity level of each district. The results indicate that Tanah Luas District has the highest vegetable productivity with a Net Flow value of 3.0608, while Tanah Pasir District ranks the lowest with a value of -2.2172. Based on these findings, the study provides recommendations for local governments and farming groups to focus on improving productivity in the lower-ranked districts. In conclusion, the PROMETHEE II method provides accurate and objective results in analyzing vegetable productivity across various districts and can serve as a decision-making tool for developing strategies to improve vegetable farming in North Aceh. Thus, the findings of this research are expected to be a reference for planning strategies to enhance agricultural productivity in North Aceh.

Keywords: *PROMETHEE* II, Vegetable Productivity, Decision Support System, District, North Aceh, *Net Flow*

1. PENDAHULUAN

Sebagian besar masyarakat Indonesia bergantung pada pertanian, yang merupakan salah satu sektor penting di negara ini [1]. Banyak jenis pertanian, termasuk pertanian sayuran. Permintaan sayuran terus meningkat, dimana konsumen harus mengeluarkan lebih banyak uang untuk sayuran setiap bulan, yang berarti bahwa harga rata-rata sayuran per kilogram yang mampu mereka beli semakin mahal. Hal ini meningkatkan produksi sayuran secara kuantitas dan kualitas [2].

Dengan adanya kelompok tani pada setiap Kecamatan di Aceh Utara, diharapkan mampu memenuhi kebutuhan masyarakat setempat. Kacang panjang, cabai, tomat, besar, dan lainnya adalah salah satu sayur yang sering ditanam. Maka dari itu, perlu dilihat kecamatan mana yang produktivitas sayurannya unggul. Hal ini dilakukan untuk memenuhi permintaan konsumen yang terus meningkat karena kesejahteraan hidup dan populasi yang meningkat.

Maka dari itu dibutuhkan sistem pendukung pengambilan keputusan dengan menggunakan metode *PROMETREE* II dalam produktivitas sayuran unggulan pada Kecamatan di Aceh Utara. Sistem Pendukung Keputusan (SPK), adalah lanjutan dari sistem informasi manajemen terkomputerisasi yang dirancang untuk berinteraksi dengan penggunanya [3]. *PROMETHEE* adalah salah satu dari banyak metode untuk menentukan perangkingan atau urutan dalam analisis multikriteria. metode ini banyak digunakan karna efisien dan sederhana, dan lebih mudah digunakan daripada metode lain untuk menyelesaikan masalah multikriteria[4].

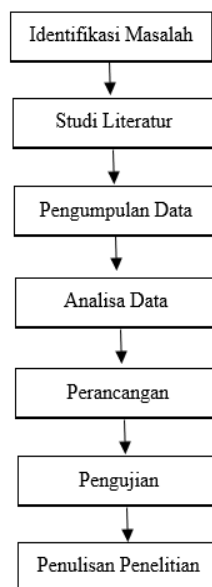
Tujuan utama melakukan penelitian ini adalah untuk melihat kabupaten mana yang menghasilkan sayuran unggulan. Begitupun sebaliknya, sehingga pemerintah setempat bisa segera menangani permasalahan tersebut. Untuk situasi ini, *Metode PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation)*, yang mengandalkan tipe preferensi dalam proses perhitungannya, berfungsi untuk menyesuaikan kebutuhan pengambil keputusan berdasarkan kriteria yang telah ditetapkan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif, yang memanfaatkan data dalam jangka waktu tertentu untuk mengatasi permasalahan [5]. Penelitian ini menggunakan data mengenai luas tanam, luas panen, luas tambah tanam, dan jumlah produksi tanaman sayur seperti cabai besar, tomat, kacang panjang, kangkung serta terung pada setiap Kecamatan di Aceh Utara selama tahun 2023. Dari keseluruhan kecamatan, hanya 23 kecamatan yang diambil sebagai sampel untuk diuji. Pemilihan sampel ini bertujuan untuk mendapatkan representasi yang cukup untuk menganalisis produktivitas sayuran di daerah masing masing kecamatan tersebut.

Penelitian sistem pendukung keputusan ini menerapkan metode *PROMETHEE* sebagai alat analisis untuk membantu proses pemeringkatan dan pengambilan keputusan. Metode ini dipilih karena kemampuannya dalam mengolah berbagai kriteria dan menghasilkan peringkat berdasarkan perbandingan alternatif [6]. Dengan menggunakan *PROMETHEE*, penelitian ini dapat memberikan hasil yang lebih objektif dan akurat dalam menentukan kecamatan dengan produktivitas sayuran terbaik hingga terendah, sehingga dapat dijadikan acuan dalam merumuskan strategi peningkatan produksi di wilayah-wilayah yang membutuhkan perhatian lebih.



Gambar 2.1 Gambar Tahapan Penelitian

Gambar 2.1 merupakan gambar tahapan penelitian. Berikut adalah penjelasannya :

1. Identifikasi masalah. Penelitian ini berfokus pada ketidakpastian dalam menentukan kecamatan dengan produktivitas sayuran tertinggi dan terendah, yang mengakibatkan kesulitan dalam mempertahankan hasil yang baik dan meningkatkan produktivitas di area yang kurang optimal.
2. Studi literatur. Pada tahap ini, dilakukan kajian literatur untuk mengumpulkan informasi atau data yang relevan dengan permasalahan yang sedang diteliti. Kajian literatur ini mencakup berbagai sumber, seperti buku, jurnal, artikel ilmiah, dan referensi lainnya yang berkaitan dengan isu yang ingin dibahas dalam penelitian ini.

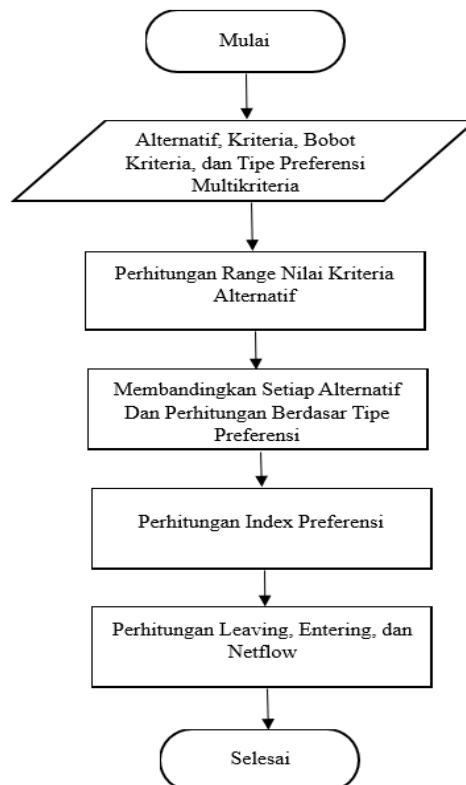
3. Pengumpulan data. Data dalam penelitian ini diperoleh dari Kelompok Wanita Tani (KWT) Meugah Raya yang terletak di Glumpang Sulu Timur, yang menyediakan data terkait produktivitas sayuran yang mendukung pelaksanaan penelitian ini.
4. Analisa data. Pada langkah ini, data yang telah dikumpulkan akan dianalisis untuk menentukan kelayakan penggunaannya sebelum merancang dan membangun sistem.
5. Perancangan. Melibatkan perencanaan sebuah sistem yang efektif, yang mencakup langkah-langkah operasional dalam pengolahan data serta prosedur yang diperlukan untuk mendukung kelancaran fungsi sistem tersebut.
6. Pengujian. Pengujian hasil perhitungan dari sistem dibandingkan dengan perhitungan manual guna memastikan ketepatan hasil yang diperoleh yang sama sama menerapkan metode *PROMETHEE*. Langkah ini bertujuan untuk memeriksa akurasi sistem dalam menghasilkan peringkat produktivitas sayuran, sehingga peneliti dapat memastikan bahwa sistem berfungsi dengan baik.
7. Penulisan penelitian. Tahap terakhir dalam penelitian ini adalah penulisan laporan yang mencakup kesimpulan komprehensif dari seluruh proses penelitian. Laporan ini juga disusun sebagai bentuk pertanggungjawaban terhadap hasil penelitian yang telah dilaksanakan.

2.2 Metode *PROMETHEE*

Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) adalah salah satu metode untuk menentukan peringkat dalam konteks *Multi Criteria Decision Making (MCDM)* [7]. Metode *PROMETHEE* digunakan untuk menentukan urutan atau prioritas dalam analisis multikriteria. Kelebihan utama dari metode ini terletak pada kesederhanaan, kejelasan, dan kestabilannya [8]. Metode ini didasarkan pada prinsip dominasi kriteria, di mana nilai-nilai yang digunakan untuk membentuk hubungan outranking menjadi dasar dalam analisisnya [9].

2.3 Tahapan Metode *PROMETHEE*

Untuk mendapatkan hasil seleksi dengan metode *PROMETHEE*, diperlukan langkah-langkah yang harus diikuti oleh pengambil keputusan.



Gambar 2.2 Skema Sistem *PROMETHEE*

Gambar 2.2 merupakan skema sistem dari metode *PROMETHEE*. Adapun maksud dari gambar tersebut dapat dilihat pada penjelasan di bawah ini :

1. Menentukan beberapa alternatif
 Alternatif dalam konteks ini merujuk pada objek yang akan dipilih atau diseleksi. Dalam proses seleksi menggunakan metode *PROMETHEE*, diperlukan penetapan beberapa objek yang akan dibandingkan (setidaknya dua objek), di mana setiap objek akan dibandingkan satu sama lain [10]
2. Menentukan beberapa kriteria
 Setelah menentukan objek yang akan diseleksi, proses perhitungan dengan metode *PROMETHEE* juga memerlukan penetapan beberapa kriteria. Kriteria tersebut berfungsi sebagai syarat atau acuan dalam proses seleksi [11]
3. Menentukan bobot kriteria
 Dalam menentukan kriteria, pengambil keputusan perlu menetapkan bobot untuk setiap kriteria. Bobot tersebut bisa memiliki nilai yang sama ataupun berbeda antar kriteria [12].
4. Menentukan tipe preferensi
 Setiap kriteria yang paling sesuai ditentukan berdasarkan data dan pertimbangan dari pengambil keputusan. Terdapat enam jenis preferensi yang dapat digunakan, yaitu Usual, Quasi, Linear, Level, Linear Quasi, dan Gaussian [13].
5. Membandingkan setiap alternatif dan perhitungan sesuai tipe preferensi.
6. Menghitung nilai preferensi, dengan rumus :

$$H(d) = \begin{cases} 0, & \text{jika } d = 0 \dots\dots\dots 1 \\ 1, & \text{jika } d \neq 0 \end{cases}$$

7. Perhitungan *Leaving Flow*, *Entering Flow*, dan *Net Flow*
 - a. Nilai *Leaving Flow* adalah *invers* dari nilai *Entering Flow*. Nilai ini mencakup total elemen yang bergerak menjauh dari suatu *node*. Dengan demikian, *Leaving Flow* dapat diartikan sebagai nilai negatif yang diberikan kepada objek seleksi yang memiliki arah menjauh dari *node* tersebut [14] . Berikut rumus untuk menghitung *Leaving Flow* :

$$\varphi^+(a) = \frac{1}{n-1} \sum_{x \in A} \varphi(a, x) \dots\dots\dots 2$$

- b. Nilai *Entering Flow*, yang juga dikenal sebagai *Negative Outranking Flow*, menunjukkan bahwa semakin kecil nilainya, semakin dominan suatu alternatif dibandingkan dengan alternatif lainnya [15]. Berikut rumus menghitung *Entering Flow*:

$$\varphi^-(a) = \frac{1}{n-1} \sum_{x \in A} \varphi(a, x) \dots\dots\dots 3$$

- c. Nilai *Net Flow* merupakan penilaian yang komprehensif. Dalam hal ini, penilaian tersebut diperoleh dengan mengurangi nilai *Leaving Flow* dari nilai *Entering Flow*. Dengan kata lain, Semakin tinggi nilai *Entering Flow* dan semakin rendah nilai *Leaving Flow*, semakin besar kemungkinan suatu alternatif untuk dipilih. Jika alternatif memiliki nilai *Net Flow* yang lebih besar, maka akan menempati peringkat yang lebih tinggi [16]. Berikut rumus untuk menghitung *Net Flow* :

$$\varphi(a) = \varphi^+(a) - \varphi^-(a) \dots\dots\dots 4$$

3 HASIL DAN PEMBAHASAN

Bagian ini berisi perhitungan dan implementasi metode *PROMETHEE* yang diperlukan untuk proses perancangan dalam produktivitas sayuran unggulan di beberapa Kecamatan di Aceh Utara. Pada perhitungan ini, metode *PROMETHEE* diimplementasikan menggunakan bahasa pemrograman *Python*.

Dataset berhasil terbaca. Berikut adalah data awal:

No	Kecamatan	Cabai_Besar_Luas_Tanam_Ha	\
0	1	Baktiya	21
1	2	Banda Baro	15
2	3	Cot Girek	23
3	4	Dewantara	12
4	5	Geureungdong Pase	11

	Cabai_Besar_Luas_Panen_Ha	Cabai_Besar_Luas_Tambah_Tanam_Ha	\
0	5	8	
1	5	5	
2	15	14	
3	7	7	
4	3	3	

	Cabai_Besar_Produksi_Kw	Tomat_Buah_Luas_Tanam_Ha	\
0	570	6	
1	196	6	
2	996	4	
3	490	6	
4	380	9	

	Tomat_Buah_Luas_Panen_Ha	Tomat_Buah_Luas_Tambah_Tanam_Ha	\
0	5	5	
1	5	5	
2	3	3	
3	4	4	
4	5	5	

Gambar 3.1 Gambar Data Awal

Gambar 3.1 merupakan sebagian data yang di tampilkan dari dataset. Data tersebut yang nantinya akan di pakai untuk melakukan perhitungan sayuran unggulan pada kecamatan yang ada di Aceh Utara.

```
alternatives = data['Kecamatan'].values

criteria = [
    'Cabai_Besar_Luas_Tanam_Ha', 'Cabai_Besar_Luas_Panen_Ha', 'Cabai_Besar_Luas_Tambah_Tanam_Ha', 'Cabai_Besar_Produksi_Kw',
    'Tomat_Buah_Luas_Tanam_Ha', 'Tomat_Buah_Luas_Panen_Ha', 'Tomat_Buah_Luas_Tambah_Tanam_Ha', 'Tomat_Buah_Produksi_Kw',
    'Kacang_Panjang_Luas_Tanam_Ha', 'Kacang_Panjang_Luas_Panen_Ha', 'Kacang_Panjang_Luas_Tambah_Tanam_Ha', 'Kacang_Panjang_Produksi_Kw',
    'Kangkung_Luas_Tanam_Ha', 'Kangkung_Luas_Panen_Ha', 'Kangkung_Luas_Tambah_Tanam_Ha', 'Kangkung_Produksi_Kw',
    'Terung_Luas_Tanam_Ha', 'Terung_Luas_Panen_Ha', 'Terung_Luas_Tambah_Tanam_Ha', 'Terung_Produksi_Kw'
]
```

Gambar 3.2 Gambar Penentuan Alternatif dan Kriteria

Gambar 3.2 merupakan tahap penentuan Alternatif dan Kriteria, yang mana setiap kecamatan yang ada pada dataset di tetapkan sebagai Alternatif yang nantinya akan di rangkingkan. Sementara, untuk kriteria perhitungannya di peroleh dari luas tanam, luas panen, luas tambah tanam, dan produksi dari ke 5 sayuran yang di tanam pada masing masing kecamatan.

```
[ ] for crit in criteria:
    min_value = data[crit].min()
    max_value = data[crit].max()
    data[crit] = (data[crit] - min_value) / (max_value - min_value)
```

Gambar 3.3 Normalisasi Data

Pada gambar 3.3 dilakukan proses normalisasi data pada nilai kriteria yang ada didalam dataset. Normalisasi data bertujuan untuk menyamakan skala berbagai kriteria agar dapat dibandingkan secara langsung sehingga dapat dievaluasi secara adil saat proses perhitungan selanjutnya.

```
def preference_function(a, b, p, q):
    diff = a - b
    if diff >= p:
        return 1
    elif q <= diff < p:
        return (diff - q) / (p - q)
    else:
        return 0

p = 0.1 # Threshold p setelah normalisasi
q = 0.05
```

Gambar 3.4 Menghitung Preferensi

Gambar 3.4 merupakan langkah untuk membandingkan dua alternatif, seperti kecamatan, berdasarkan berbagai kriteria yang telah ditetapkan. Preferensi ini menunjukkan seberapa unggul satu alternatif dibandingkan yang lain pada masing-masing kriteria, seperti luas tanam atau tingkat produksi. Hasil dari preferensi ini kemudian digunakan untuk menghitung "leaving flow" dan "entering flow," yang pada akhirnya akan menentukan peringkat keseluruhan dari alternatif yang ada.

```
[ ] leaving_flow = np.sum(preference_matrix, axis=1) / (n - 1)
    entering_flow = np.sum(preference_matrix, axis=0) / (n - 1)
```

Gambar 3.5 Menghitung Leaving dan Entering Flow

Gambar 3.5 ialah perhitungan *Leaving flow* dan *Entering flow* merupakan dua aspek penting dalam metode *PROMETHEE* yang berfungsi untuk menentukan urutan alternatif. Selisih antara *leaving flow* dan *entering flow* digunakan untuk menghitung nilai *net flow*, yang kemudian dijadikan acuan dalam penentuan peringkat akhir.

```
[ ] net_flow = leaving_flow - entering_flow
```

Gambar 3.6 Menghitung Netflow

Gambar 3.6 merupakan aspek paling penting dalam metode *Promethee*, *Net flow* menggambarkan seberapa unggul suatu alternatif (kecamatan) secara keseluruhan dibandingkan dengan alternatif lainnya. Ketika nilai *net flow* positif, alternatif tersebut lebih diutamakan dibandingkan yang lain, sementara nilai negatif menandakan sebaliknya. *Net flow* inilah yang menjadi dasar dalam menentukan urutan akhir dari semua alternatif dalam proses pengambilan keputusan.

Berikut adalah hasil sebagian tabel peringkat yang disusun berdasarkan nilai *Netflow*, yang mencakup 22 peringkat dari yang terbaik hingga terburuk, berdasarkan produktivitas sayuran di setiap kecamatan.

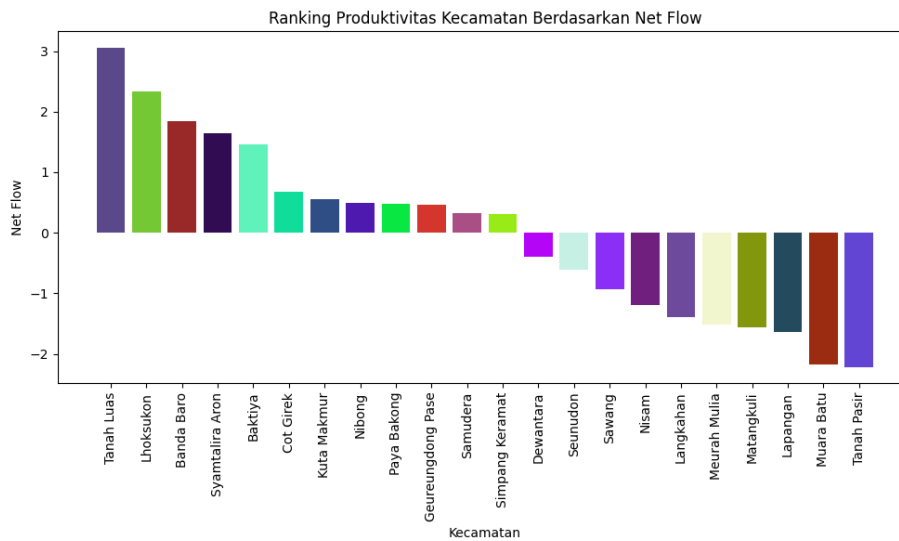
Tabel 3.1 Hasil Perangkingan Dari Nilai *NetFlow*

Ranking	Kecamatan	Net Flow
1	Tanah Luas	3.0608
2	Lhoksukon	2.3333
3	Banda Baru	1.8429

Tabel 3.2 Hasil Perangkingan Dari Nilai *NetFlow* (Lanjutan)

Ranking	Kecamatan	Net Flow
4	Syamtalira Aron	1.6380
-	-	-
19	Matangkuli	-1.5673
20	Lapangan	-1.6309
21	Muara Batu	-2.1747
22	Tanah Pasir	-2.2172

Berdasarkan tabel di atas, nilai *Net Flow* yang diperoleh menjadi faktor penentu dalam proses perankingan kecamatan. Hasilnya menunjukkan bahwa Kecamatan Tanah Luas memiliki produktivitas sayuran yang paling tinggi, sementara Kecamatan Tanah Pasir berada di peringkat terbawah dengan produktivitas yang paling rendah.



Gambar 3.7 Grafik Perankingan

Pada Gambar 3.7 ditampilkan seluruh ranking berdasarkan nilai *Net Flow*. Kategorisasi dapat dilakukan sebagai berikut: kecamatan dengan nilai *Net Flow* di atas 1 dikategorikan sebagai sangat baik, sedangkan nilai antara 0 hingga 1 dianggap baik. Nilai antara -1 hingga 0 menunjukkan kategori tidak baik, sementara kecamatan dengan nilai *Net Flow* di bawah -1 dikategorikan sebagai sangat tidak baik. Pendekatan ini memungkinkan analisis yang lebih jelas mengenai tingkat produktivitas sayuran di setiap kecamatan.

4. DISKUSI

Hasil dari penelitian ini menunjukkan bahwa metode *PROMETHEE II* sangat efektif dalam menentukan peringkat produktivitas sayuran di berbagai kecamatan di Aceh Utara berdasarkan nilai *Net Flow*. Kecamatan Tanah Luas menempati peringkat tertinggi dengan nilai *Net Flow* sebesar 3.0608, sedangkan Kecamatan Tanah Pasir berada pada posisi terendah dengan nilai -2.2172. Temuan ini memberikan wawasan yang penting bagi pemerintah daerah dan kelompok tani untuk memprioritaskan peningkatan produktivitas di kecamatan dengan peringkat rendah.

Hasil penelitian ini konsisten dengan beberapa penelitian sebelumnya yang juga menggunakan metode *PROMETHEE* dalam konteks penilaian multikriteria di bidang pertanian. Misalnya, penelitian oleh Kurniawan & Rahmadani (2020) yang menggunakan metode *PROMETHEE* untuk pemilihan bibit cabai menunjukkan bahwa metode ini mampu menghasilkan keputusan yang lebih objektif dibandingkan dengan metode lain seperti AHP (*Analytic Hierarchy Process*). Di bidang pertanian sayuran, penelitian Mauliddiyah (2021) juga menemukan bahwa metode *PROMETHEE* sangat bermanfaat dalam menganalisis lahan pertanian yang optimal di kota Batu.

Namun, penelitian ini memiliki keterbatasan yang perlu diperhatikan. Salah satunya adalah keterbatasan data yang hanya mencakup 23 kecamatan, yang mungkin tidak mencakup seluruh keragaman kondisi pertanian di Aceh Utara secara menyeluruh. Selain itu, hanya empat kriteria yang digunakan, yakni luas tanam, luas panen, luas tambah tanam, dan produksi. Penambahan kriteria lain, seperti kualitas tanah atau ketersediaan air, mungkin akan memberikan hasil yang lebih komprehensif.

Secara keseluruhan, penelitian ini memberikan kontribusi signifikan dalam bidang produktivitas pertanian sayuran dan menunjukkan bagaimana metode *PROMETHEE II* dapat diimplementasikan secara efektif dalam membantu pengambilan keputusan yang kompleks. Hasil penelitian ini dapat menjadi dasar bagi pengambilan kebijakan dan strategi untuk meningkatkan produktivitas pertanian di Aceh Utara, terutama di kecamatan dengan peringkat produktivitas yang rendah.

5. KESIMPULAN

Penelitian ini menyimpulkan bahwa metode *PROMETHEE* efektif dalam menghasilkan peringkat produktivitas kecamatan berdasarkan nilai *Net Flow*. Dari hasil analisis, Kecamatan Tanah Luas berada di posisi teratas dengan nilai *Net Flow* sebesar 3.0608, sedangkan Kecamatan Tanah Pasir menduduki posisi terbawah dengan nilai -2.2172. Kecamatan dengan produktivitas yang baik sebaiknya menjaga performanya, sedangkan kecamatan dengan hasil yang kurang memuaskan perlu berusaha untuk meningkatkan produktivitas pertanian. Dengan

menggunakan metode *PROMETHEE*, petani dan pemerintah daerah dapat lebih memahami tingkat produktivitas sayuran, sehingga dapat mengambil langkah-langkah yang tepat untuk meningkatkan hasil pertanian di daerah mereka.

DAFTAR PUSTAKA

- [1] A. Ulya, M. Muliadi, R. Herteno, A. Farmadi, And F. Abadi, "Sistem Pendukung Keputusan Pemilihan Bibit Padi Berkualitas Pada Lahan Rawa Menggunakan Metode *Dematel* Dan Mfep," *Sebatik*, Vol. 28, No. 1, Pp. 85–92, 2024, Doi: 10.46984/Sebatik.V28i1.2291.
- [2] N. L. Mauliddiyah, "Mauliddiyah, Nurul L, 'Implementasi Metode Ahp-Vikor Untuk Pemilihan Lahan Pertanian Sayuran Di Kota Batu,'" P. 6, 2021.
- [3] E. Kurniawan And N. Rahmadani, "Perancangan Sistem Pendukung Keputusan Pemilihan Bibit Cabai Unggul Menggunakan Metode Fmcdm Tanaman Cabai Merupakan Salah Satu Sayuran Buah Yang Memiliki Peluang Bisnis Yang Baik . Besarnya Kebutuhan Dalam Negeri Maupun Luar Negeri Menjadikan Cabai Sebag," Pp. 957–965, 2020.
- [4] E. Novida And H. Sunandar, "Sistem Pendukung Keputusan Pemilihan Produk Lensa Kacamata Menggunakan Metode *Promethee* Ii," *Pelita Inform. Inf. Dan Inform.*, Vol. 6, No. 3, Pp. 325–332, 2018.
- [5] H. Hatta And N. Nursanty, "Penentuan Produktivitas Daerah Kabupaten Terbaik Sebagai Penghasil Tanaman Sayuran Cabai Di Provinsi Sumatera Selatan Menggunakan Diagram Kartesius," *Publ. Penelit. Terap. Dan Kebijakan*, Vol. 3, No. 1, Pp. 26–33, 2020, Doi: 10.46774/Pptk.V12i1.103.
- [6] N. R. Muntiari, S. Sunardi, And A. Fadlil, "Sistem Penentuan Penginapan Dengan Metode *Promethee*," *J. Ilm. Mandala Educ.*, Vol. 6, No. 1, Pp. 12–19, 2020, Doi: 10.58258/Jime.V6i1.1098.
- [7] R. O. Siregar, D. Irmayani, And M. Masrizal, "Penerapan Metode *Promethee* Dalam Sistem Pendukung Keputusan Penetapan Penerima Kartu Indonesia Sehat (Kis)," *J. Media Inform. Budidarma*, Vol. 5, No. 2, P. 739, 2021, Doi: 10.30865/Mib.V5i2.2948.
- [8] H. Geke, Y. R. Kaesmetan, D. Prasetyo, And M. Saitakela, "Identifikasi Penyakit Pada Tanaman Kakao Menggunakan *Promethee*," *Pros. Semin. Nas. Sist. Inf. Dan Teknol.*, Pp. 215–219, 2022, [Online]. Available: <https://www.seminar.laii.or.id/index.php/sisfotek/article/view/373/308>
- [9] M. I. Suhada, I. S. Damanik, I. S. Saragih, S. R. Andani, And H. Siahaan, "Sistem Pendukung Keputusan Menentukan Kenaikan Jabatan Pegawai Kantor Kejaksaan Negeri Pematangsiantar," *Kaji. Ilm. Inform. Dan Komput.*, Vol. 1, No. 5, Pp. 199–203, 2021.
- [10] F. Yasari Br Surbakti, A. Fauzi, And S. Ramadani, "Pengambilan Keputusan Penentuan Target Promosi Kampus Stmik Kaputama Dengan Metode *Promethee*," *J. Jisiikom (Jurnal Inov. Sist. Inf. Ilmu Komputer)*, Vol. 2, No. 1, Pp. 3025–4868, 2024, [Online]. Available: <https://jisiikom.org/index.php/journal/article/view/18>
- [11] M. Rosmiati, "Implementasi Metode *Promethee* Dalam Menentukan Kenaikan Jabatan Pada Nuansa Agung Garment," *Multinetics*, Vol. 8, No. 2, Pp. 160–167, 2023, Doi: 10.32722/Multinetics.V8i2.5140.
- [12] U. Juhardi, K. Khairullah, Y. Apridiansyah, And A. Kusnandi, "Implementasi Metode *Promethee* Seleksi Penerima Bantuan Langsung Tunai," *J-Sisko Tech (Jurnal Teknol. Sist. Inf. Dan Sist. Komput. Tgd)*, Vol. 6, No. 1, P. 76, 2023, Doi: 10.53513/Jsk.V6i1.7389.
- [13] M. Arif And N. Hafidz, "Sistem Pendukung Keputusan Pemilihan Supplier Menggunakan Metode *Promethee* (Studi Kasus : Cathead Apparel)," Vol. 9, No. 3, Pp. 142–154, 2021.
- [14] S. R. Handayani And B. Noranita, "Penerapan Metode *Promethee* Dalam Menentukan Prioritas Penerima Kredit," *J. Masy. Inform.*, Vol. 9, No. 2, Pp. 1–9, 2018, Doi: 10.14710/Jmasif.9.2.31485.
- [15] A. Abdulrajak *Et Al.*, "Dss For Selection Types Food Plants Based On Environmental Conditions In Tidore City Using *Promethee* Methods," *Jiko (Jurnal Inform. Dan Komputer)*, Vol. 3, No. 2, Pp. 87–91, 2020, Doi: 10.33387/Jiko.
- [16] S. Saifulloh, "Analisis *Promethee* II Sebagai Pendukung Keputusan Pemilihan Media Sosial," *J. Tek. Inform. Dan Sist. Inf.*, Vol. 6, No. 3, Pp. 515–520, 2020, Doi: 10.28932/Jutisi.V6i3.2956.

SENASTIKA Universitas Malikussaleh

Identifikasi Siswa Kelas Unggulan SMK Padang Panjang Dengan Menggunakan Algoritma k-Means Clustering

M Wanda*¹, Azna Alaiya², Azizi Jabbar³, Alayda Zaielamy Putri⁴
Munirul Ula⁵

^{1,2,3,4} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹wanda.210170041@mhs.unimal.ac.id,
²azna.210170050@mhs.unimal.ac.id,³azizi.210170070@mhs.unimal.ac.id,
⁴alayda.210170047@mhs.unimal.ac.id,⁵munirulula@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam pengelompokan siswa berdasarkan nilai semester 1 dan semester 2 untuk menentukan kelas unggulan di SMK Padang Panjang. Data yang digunakan terdiri dari 388 siswa, di mana proses clustering dilakukan dengan menghitung jarak antara data dan centroid menggunakan rumus Euclidean. Algoritma K-Means dipilih karena kemampuannya yang sederhana dan efektif untuk mengelola data dalam jumlah besar. Setelah melalui empat iterasi, hasil clustering mencapai kestabilan pada iterasi ke-3 dan ke-4, dengan pembagian cluster dan nilai centroid yang tidak mengalami perubahan. Hal ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan siswa secara optimal berdasarkan kriteria yang telah ditentukan, sehingga dapat mendukung pengambilan keputusan terkait penentuan kelas unggulan.

Keywords: *K-Means, clustering, data mining, pengelompokan siswa*

1. PENDAHULUAN

Metode pembelajaran yang bertujuan untuk meningkatkan pengetahuan, sikap, dan keterampilan seseorang dikenal sebagai pendidikan. Sekolah telah berubah menjadi sarana berlangsungnya pendidikan secara langsung, di mana guru dan siswa bertemu[1]. Kualitas siswa dalam belajar di sekolah sangat beragam, yang menyebabkan tingkat pemahaman siswa yang berbeda. Sekolah harus memperhatikan hal ini, terutama guru sebagai pendidik dan pendidik siswa. Salah satunya dengan membuat semua siswa memiliki kesempatan untuk berprestasi dalam kelompok belajar yang efektif. Dalam belajar kelompok, lebih dari satu orang saling melengkapi dan berbagi ide[2]. Dimungkinkan untuk mengurangi tingkat kesulitan yang dihadapi siswa saat belajar di sekolah dengan menerapkan metode belajar kelompok[3].

Tingginya prestasi siswa dan rendahnya jumlah siswa yang tidak berprestasi mencerminkan keunggulan sektor Pendidikan[4]. Saat ini, sektor pendidikan diharapkan mampu bersaing dengan memanfaatkan kemajuan teknologi informasi (TI), yang dapat meningkatkan daya saing serta mendukung operasional sehari-hari dan pengambilan keputusan strategis[5]. Secara umum, keberhasilan siswa dievaluasi berdasarkan penilaian terhadap pelajaran teoretis dan praktis, serta kehadiran dan ketidakhadiran yang mempengaruhi nilai mereka[6].

Jumlah data siswa terus meningkat setiap tahunnya[7]. Hal ini menyebabkan penumpukan data yang belum diolah dengan baik, yang seharusnya dapat digunakan untuk mengungkap pengetahuan dan informasi baru berdasarkan pola yang terbentuk dari akumulasi data tersebut[8]. Peningkatan volume pemrosesan data ini membutuhkan penerapan strategi dan metode agar data tersebut dapat diubah menjadi informasi dan pengetahuan yang berguna bagi pendidik dalam pembuatan kebijakan[9]. Hal ini menjadi tantangan bagi guru dan wali kelas, yaitu bagaimana cara menentukan siswa kelas yang unggul.

Clustering adalah teknik Data Mining. Salah satu metode analisis data yang dapat digunakan untuk memecahkan masalah pengelompokan data adalah metode K-means, yang dapat mengelola sejumlah besar data dalam waktu yang relatif cepat dan efisien[10].

Tujuan dari penelitian ini yaitu untuk mengelompokkan siswa berdasarkan kriteria performa akademik, Menggunakan algoritma k-Means untuk mengelompokkan siswa SMK berdasarkan nilai akademik mereka, sehingga dapat diidentifikasi kelompok siswa yang berpotensi untuk ditempatkan di kelas unggulan.

2. METODE PENELITIAN

Peneliti melakukan berbagai proses, dan setiap proses dapat membantu penelitian ini. Dalam penelitian, tahapan penelitian menjelaskan proses yang diambil mulai dari menemukan masalah hingga menemukan solusinya. Tahap penelitiannya sebagai berikut :

1. Metode Pengumpulan data

Dilakukan dengan cara mengambil data melalui kaggle, Dengan jumlah keseluruhan data 388 data siswa, nilai – nilai yang di ambil dari siswa tersebut adalah nilai semester 1 dan semester 2. Link kaggle: <https://www.kaggle.com/datasets/laychristian/data-siswa-smk>.

2. Identifikasi Masalah

Tahap mengidentifikasi masalah yang terjadi, yaitu kurangnya informasi yang tepat dan rinci tentang menggabungkan data nilai siswa berdasarkan hasil penilaian proses pembelajaran siswa yang dilakukan setiap semester.

3. Pendekatan

Metode clustering mengumpulkan data nilai siswa berdasarkan hasil penilaian proses pembelajaran untuk menentukan kelas unggulan siswa berdasarkan nilai semester, yang berdampak pada data evaluasi siswa.

4. Pengolahan Data

Pengolahan data terdiri dari menyiapkan data nilai siswa berdasarkan nilai rata-rata dari nilai semester 1 dan 2 yang akan digunakan, menentukan jumlah cluster yang akan digunakan dan menghitung masing-masing kriteria cluster berdasarkan nilai rata-rata semester 1 dan 2.

5. Clustering

Pengelompokan objek berdasarkan hubungan dan kemiripan dalam kelas kelas yang sama dikenal sebagai clustering. Mengelompokkan keunggulan siswa berdasarkan nilai rata-rata semester 1 dan 2, maka implementasi algoritma k-Means Clustering akan difokuskan pada variabel nilai akademik siswa dari kedua semester.

6. K- Means

Algoritma K-means sederhana dan efektif untuk kumpulan data yang besar jika dibandingkan dengan pengelompokan hierarkis, algoritma ini digunakan untuk inisialisasi parameter. Berikut ini adalah Langkah-langkah algoritma K-Means, sebagai berikut :

- 1) Tentukan jumlah cluster yang mau dibentuk.
- 2) Tentukan pusat cluster sebanyak jumlah cluster yang mau dibentuk secara acak.
- 3) Hitung jarak masing-masing data terhadap masing-masing pusat cluster.

Rumus untuk menghitung jarak menggunakan rumus Euclidean Distance

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- 4) Tentukan cluster yang terbentuk, lalu tentukan pusat cluster baru, dan hitung Kembali jaraknya.
- 5) Mengulangi langkah sebelumnya sampai hasil pengelompokkan tidak mengalami perubahan [10].

3. HASIL DAN PEMBAHASAN

Untuk menentukan siswa kelas unggulan dengan mengambil 388 data siswa sesuai dengan kriteria perhitungan. Kriteria yang digunakan untuk melakukan perhitungan K-Means ini ditentukan sesuai dengan kebutuhan yaitu kriteria yang diambil dari nilai semester 1 dan semester 2. Adapun daftar kriteria semester 1 dan 2 yang digunakan dapat dilihat dari tabel 1 berikut :

Tabel 1. Kriteria yang digunakan

No	Semester	Inisialisasi
1.	Semester 1	S1
2.	Semester 2	S2

Data awal sebelum dilakukan perhitungan untuk menentukan siswa kelas unggulan dengan penilaian semester 1 dan 2 yang menjadi kriteria, data -data tersebut dapat dilihat pada tabel 2 berikut :

Tabel 2. Sampel data nilai siswa

Nama Siswa	S1	S2
Siswa 1	64,6	63,6
Siswa 2	65,8	65
Siswa 3	62,4	72,2
Siswa 4	72,4	72,8
Siswa 5	64,6	65,2
Siswa 6	65	64,2
Siswa 7	62,6	64,4
Siswa 8	65,6	66,8
Siswa 9	64,6	66,2
Siswa 10	78,6	81,1
...
Siswa 379	70,4	72,6

Siswa 380	67,4	73,6
Siswa 381	75,4	76,8
Siswa 382	57,2	68,2
Siswa 383	73,4	74
Siswa 384	72,2	74,4
Siswa 385	71,4	77
Siswa 386	67,4	74,8
Siswa 387	72,2	74,2
Siswa 388	74	74,8

Selanjutnya menentukan jumlah cluster, jumlah cluster yang mau dibentuk yaitu 2 cluster, dimana pusat clusternya (Centroid) yaitu pada siswa 1 (C1) dan siswa 2 (C2). Pusat awal cluster terdapat pada tabel 3 berikut:

Tabel 3. Pusat awal cluster

Centroid	S1	S2
C1	64,6	63,6
C2	65,8	65

Hitung jarak setiap data yang ada terhadap setiap pusat cluster. Misalkan untuk perhitungan jarak data siswa pertama dengan pusat cluster pertama dengan rumus euclidian distance.

Perhitungan jarak data pertama dengan pusat cluster pertama adalah :

$$d_{(Siswa1. C1)} = \sqrt{(64,6 - 64,6)^2 + (63,6 - 63,6)^2} = \sqrt{0 + 0} = 0$$

Perhitungan jarak data pertama dengan pusat cluster kedua adalah :

$$d_{(Siswa1. C2)} = \sqrt{(64,6 - 65,8)^2 + (63,6 - 65,0)^2} = 3,4 = 1,84 = \sqrt{3,4} = 1,8439$$

Berikut adalah hasil perhitungan lengkap dari perhitungan clustering:

Tabel 4. Hasil Perhitungan Iterasi Ke-1

Nama Siswa	S1	S2	C1	C2	Jarak Terdekat
Siswa 1	64,6	63,6	0,000	1,844	C1
Siswa 2	65,8	65	1,844	0,000	C2
Siswa 3	62,4	72,2	8,877	7,962	C2
Siswa 4	72,4	72,8	12,062	10,218	C2
Siswa 5	64,6	65,2	1,600	1,217	C2
Siswa 6	65	64,2	0,721	1,131	C1
Siswa 7	62,6	64,4	2,154	3,256	C1
Siswa 8	65,6	66,8	3,353	1,811	C2
Siswa 9	64,6	66,2	2,600	1,697	C2
Siswa 10	78,6	81,1	22,411	20,568	C2
...
Siswa 379	70,4	72,6	10,707	8,884	C2
Siswa 380	67,4	73,6	10,385	8,748	C2
Siswa 381	75,4	76,8	17,055	15,212	C2
Siswa 382	57,2	68,2	8,713	9,176	C1
Siswa 383	73,4	74	13,624	11,780	C2
Siswa 384	72,2	74,4	13,206	11,372	C2
Siswa 385	71,4	77	15,027	13,242	C2
Siswa 386	67,4	74,8	11,545	9,930	C2
Siswa 387	72,2	74,2	13,043	11,207	C2
Siswa 388	74	74,8	14,622	12,778	C2

Berasarkan tabel 4 di atas, data kemudian dikelompokkan menjadi dua cluster dengan rincian sebagai berikut.

Tabel 5. Pengelompokkan Iterasi Ke-1

Cluster	Pengelompokkan	Jumlah Data
C1	1,6,7,25,40,41,49,72,77,97,100,104,116,117,121,124,146,149,172,212,217,225,258,260,261,288,293,338,371,382	30
C2	2,3,4,5,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,26,27,28,29,30,31,32,33,34,35,36,37,38,39,42,43,44,45,46,47,48,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,73,74,75,76,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,98,99,101,102,103,105,106,107,108,109,110,111,112,113,114,	358

4 SENASTIKA 2024, Jurusan Informatika Universitas Malikussaleh

115,118,119,20,122,123,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,147,148,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,213,214,215,216,218,219,220,221,222,223,224,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,256,257,259,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,178,279,280,281,282,283,284,285,286,287,289,290,291,292,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,372,373,374,375,376,377,378,379,380,381,383,384,385,386,387,388
--

Berikut ini merupakan hasil rata-rata yang diperoleh pada iterasi ke-1 untuk setiap cluster.

Tabel 6. Rata-Rata Iterasi Ke-1

CLUSTER	S1	S2
C1	57,573	66,660
C2	69,973	73,185

Melanjutkan proses dengan menghitung kembali jarak data ke centroid menggunakan rumus Euclidean untuk memastikan pembaruan kelompok secara akurat (tidak ada perubahan).

Tabel 7. Hasil Perhitungan Iterasi Ke-2

Nama Siswa	S1	S2	C1	C2	Jarak Terdekat
Siswa 1	64,6	63,6	5,646	10,988	C1
Siswa 2	65,8	65	6,172	9,187	C1
Siswa 3	62,4	72,2	6,097	7,637	C1
Siswa 4	72,4	72,8	13,967	2,457	C2
Siswa 5	64,6	65,2	4,965	9,624	C1
Siswa 6	65	64,2	5,703	10,269	C1
Siswa 7	62,6	64,4	3,556	11,469	C1
Siswa 8	65,6	66,8	5,747	7,739	C2
Siswa 9	64,6	66,2	4,767	8,812	C1
Siswa 10	78,6	81,1	23,635	11,678	C2
...
Siswa 379	70,4	72,6	12,103	0,724	C2
Siswa 380	67,4	73,6	10,251	2,606	C2
Siswa 381	75,4	76,8	18,560	6,521	C2
Siswa 382	57,2	68,2	3,069	13,711	C1
Siswa 383	73,4	74	15,406	3,523	C2
Siswa 384	72,2	74,4	14,571	2,537	C2
Siswa 385	71,4	77	15,498	4,073	C2
Siswa 386	67,4	74,8	11,099	3,038	C2
Siswa 387	72,2	74,2	14,465	2,447	C2
Siswa 388	74	74,8	16,320	4,339	C2

Mengacu pada tabel 7 di atas, data selanjutnya dibagi ke dalam dua cluster dengan detail sebagai berikut.

Tabel 8. Pengelompokkan Iterasi Ke-2

Cluster	Pengelompokkan	Jumlah Data
C1	1,2,3,5,6,7,9,25,29,38,40,41,47,49,55,56,72,77,80,91,95,97,98,100,102,104,114,116,117,121,124,134,143,146,149,172,177,183,197,201,208,209,212,217,225,243,253,258,259,260,261,264,267,269,270,287,288,293,301,308,338,356,370,371,376,382	66
C2	4,8,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,26,27,28,30,31,32,33,34,35,36,37,39,42,43,44,45,46,48,50,51,52,53,54,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,73,74,75,76,78,79,81,82,83,84,85,86,87,88,89,90,92,93,94,96,98,99,101,103,105,106,107,108,109,110,111,112,113,115,118,119,120,122,123,125,126,127,128,129,130,131,132,133,135,136,137,138,139,140,141,142,144,145,147,148,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,173,174,175,176,178,179,180,181,182,184,185,186,187,188,189,190,191,192,193,194,195,196,198,199,200,202,203,204,205,206,207,210,211,213,214,215,216,218,219,220,221,222,223,224,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,244,245,246,247,248,249,250,251,252,254,255,256,257,262,263,265,266,268,269,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,289,290,291,292,294,295,296,297,298,299,300,302,303,304,305,306,307,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,339,340,341,342,343,344,345,346,347,348,3	322

49,350,351,352,353,354,355,357,358,359,360,361,362,363,364,365,366,367,368,369,,372,373,374,375,377,378,379,380,381,383,384,385,386,387,388

Hasil rata-rata untuk setiap cluster pada iterasi ke-2 disajikan sebagai berikut.

Tabel 9. Rata-Rata Iterasi Ke-2

CLUSTER	S1	S2
C1	59,855	67,809
C2	63,763	66,270

Langkah berikutnya adalah menghitung ulang jarak antara data dan centroid menggunakan rumus Euclidean, guna memastikan pembaruan cluster dilakukan dengan tepat (tidak ada perubahan).

Tabel 10. Hasil Perhitungan Iterasi Ke-3

Nama Siswa	S1	S2	C1	C2	Jarak Terdekat
Siswa 1	64,6	63,6	6,343	2,798	C2
Siswa 2	65,8	65	6,575	2,400	C2
Siswa 3	62,4	72,2	5,075	6,085	C1
Siswa 4	72,4	72,8	13,501	10,828	C2
Siswa 5	64,6	65,2	5,415	1,358	C2
Siswa 6	65	64,2	6,285	2,411	C2
Siswa 7	62,6	64,4	4,377	2,202	C2
Siswa 8	65,6	66,8	5,833	1,912	C2
Siswa 9	64,6	66,2	5,010	0,840	C2
Siswa 10	78,6	81,1	22,953	20,947	C2
...
Siswa 379	70,4	72,6	11,582	9,172	C2
Siswa 380	67,4	73,6	9,511	8,183	C2
Siswa 381	75,4	76,8	17,958	15,694	C2
Siswa 382	57,2	68,2	2,684	6,841	C1
Siswa 383	73,4	74	14,893	12,354	C2
Siswa 384	72,2	74,4	13,994	11,717	C2
Siswa 385	71,4	77	14,757	13,170	C2
Siswa 386	67,4	74,8	10,286	9,273	C2
Siswa 387	72,2	74,2	13,901	11,579	C2
Siswa 388	74	74,8	15,778	13,325	C2

Mengacu pada tabel 10 di atas, data selanjutnya dibagi ke dalam dua cluster dengan detail sebagai berikut.

Tabel 11. Pengelompokan Iterasi Ke-3

Cluster	Pengelompokan	Jumlah Data
C1	3,47,49,72,80,97,100,102,114,116,121,134,146,149,177,208,212,217,224,243,253,256,257,258,259,260,287,293,301,308,338,371,378,382	32
C2	1,2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,48,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,73,74,75,76,77,78,79,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,98,99,101,103,104,105,106,107,108,109,110,111,112,113,115,117,118,119,120,122,123,124,125,126,127,128,129,130,131,132,133,135,136,137,138,139,140,141,142,143,144,145,147,148,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,209,210,211,213,214,215,216,218,219,220,221,222,223,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,244,245,246,247,248,249,250,251,252,254,255,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,288,289,290,291,292,294,295,296,297,298,299,300,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,372,373,374,375,376,377,379,380,381,383,384,385,386,387,388	356

6 SENASTIKA 2024, Jurusan Informatika Universitas Malikussaleh

Rata-rata hasil yang diperoleh dari iterasi ke-3 untuk masing-masing cluster ditampilkan di bawah ini.

Tabel 12. Rata-Rata Iterasi Ke-3

CLUSTER	S1	S2
C1	59,375	75,788
C2	69,881	72,401

Proses dilanjutkan dengan menghitung kembali jarak antara setiap data dan centroid menggunakan rumus Euclidean, memastikan pembagian cluster diperbarui secara akurat.

Tabel 13. Hasil Perhitungan Iterasi Ke-4

Nama Siswa	S1	S2	C1	C2	Jarak Terdekat
Siswa 1	64,6	63,6	13,261	10,264	C2
Siswa 2	65,8	65	12,556	8,452	C2
Siswa 3	62,4	72,2	4,693	7,484	C1
Siswa 4	72,4	72,8	13,363	2,550	C2
Siswa 5	64,6	65,2	11,807	8,930	C2
Siswa 6	65	64,2	12,881	9,544	C2
Siswa 7	62,6	64,4	11,836	10,818	C2
Siswa 8	65,6	66,8	10,933	7,050	C2
Siswa 9	64,6	66,2	10,919	8,145	C2
Siswa 10	78,6	81,1	19,934	12,285	C2
...
Siswa 379	70,4	72,6	11,477	0,556	C2
Siswa 380	67,4	73,6	8,318	2,755	C2
Siswa 381	75,4	76,8	16,057	7,057	C2
Siswa 382	57,2	68,2	7,894	13,359	C1
Siswa 383	73,4	74	14,139	3,865	C2
Siswa 384	72,2	74,4	12,900	3,061	C2
Siswa 385	71,4	77	12,086	4,843	C2
Siswa 386	67,4	74,8	8,086	3,451	C2
Siswa 387	72,2	74,2	12,923	2,935	C2
Siswa 388	74	74,8	14,658	4,766	C2

Mengacu pada tabel 13 di atas, data selanjutnya dibagi ke dalam dua cluster dengan detail sebagai berikut.

Tabel 14. Pengelompokan Iterasi Ke-4

Cluster	Pengelompokan	Jumlah Data
C1	3,47,49,72,80,97,100,102,114,116,121,134,146,149,177,208,212,217,224,243,253,256,257,258,259,260,287,293,301,308,338,371,378,382	32
C2	1,2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,48,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,73,74,75,76,77,78,79,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,98,99,101,103,104,105,106,107,108,109,110,111,112,113,115,117,118,119,120,122,123,124,125,126,127,128,129,130,131,132,133,135,136,137,138,139,140,141,142,143,144,145,147,148,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,209,210,211,213,214,215,216,218,219,220,221,222,223,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,244,245,246,247,248,249,250,251,252,254,255,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,288,289,290,291,292,294,295,296,297,298,299,300,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,372,373,374,375,376,377,379,380,381,383,384,385,386,387,388	356

Rata-rata hasil yang diperoleh dari iterasi ke-4 untuk masing-masing cluster ditampilkan di bawah ini.

Tabel 15. Rata-Rata Iterasi Ke-4

CLUSTER	S1	S2
C1	59,375	75,788

C2	69,881	72,401
----	--------	--------

Setelah dilakukan sebanyak 4 iterasi, terlihat bahwa pada iterasi ke-3 dan ke-4 memiliki anggota yang sama, dan nilai centroidnya tidak mengalami perubahan. Oleh karena itu, iterasi lebih lanjut tidak diperlukan karena hasil clustering telah mencapai kestabilan dan konvergen.

Siswa kelas unggulan adalah yang masuk ke dalam Cluster C2, karena cluster ini memiliki nilai rata-rata Semester 1 dan 2 yang lebih tinggi dibandingkan Cluster C1. Jumlah siswa kelas unggulan dari total 388 siswa yang dikelompokkan, 356 siswa masuk ke dalam Cluster C2, yang dapat dikategorikan sebagai siswa kelas unggulan.

5. DISKUSI

Identifikasi Siswa Kelas Unggulan SMK Padang Panjang Dengan Menggunakan Algoritma k-Means Clustering memberikan gambaran lengkap tentang proses penerapan k-Means Clustering untuk identifikasi siswa kelas unggulan di SMK Padang Panjang dan bagaimana hasilnya bisa dimanfaatkan untuk keperluan pendidikan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means berhasil digunakan untuk mengelompokkan siswa ke dalam dua cluster berdasarkan nilai semester 1 dan semester 2 dari 388 siswa. Proses clustering dilakukan dengan menghitung jarak antara data dan centroid menggunakan rumus Euclidean. Setelah melalui empat iterasi, hasil clustering mencapai kestabilan pada iterasi ke-3 dan ke-4, di mana pembagian cluster dan nilai centroid tidak mengalami perubahan. Oleh karena itu, iterasi lebih lanjut tidak diperlukan, menunjukkan bahwa algoritma K-Means telah berhasil mengelompokkan siswa secara optimal sesuai kriteria yang digunakan.

DAFTAR PUSTAKA

- [1] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [2] I. D. Setiawan and A. Triayudi, "Penerapan Data Mining Dengan Menggunakan Algoritma Clustering K-Means Untuk Pembagian Jurusan Pada Sekolah Menengah Atas," *J. Comput. Syst. Informatics*, vol. 5, no. 2, pp. 380–392, 2024, doi: 10.47065/josyc.v5i2.4970.
- [3] T. Hidayat, "Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering," *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 19–24, 2022, doi: 10.37034/jsisfotek.v4i1.115.
- [4] M. S. Said and Y. Yusti, "Penerapan Algoritma K-Means Dalam Penentuan Jurusan Siswa Sman 05 Bombana," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 5, no. 2, pp. 114–122, 2020, doi: 10.51876/simtek.v5i2.87.
- [5] I. Savitri, H. 1, S. Defit, and G. W. Nurcahyo, "Jurnal Sistim Informasi dan Teknologi <https://jsisfotek.org/index.php> Simulasi dalam Optimalisasi Pengadaan Barang menggunakan Metode K-Mean Clustering," 2021, doi: 10.37034/jsisfotek.v3i3.79.
- [6] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [7] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [8] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.
- [9] N. R. Wahyudi, Y. Rahmawati, and S. Supriyanto, "Implementasi Rapidminer Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 7, no. 2, pp. 109–118, 2024, doi: 10.36080/skanika.v7i2.3173.
- [10] R. P. Primanda, A. Alwi, and D. Mustikasari, "DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," *Komputek*, vol. 5, no. 1, p. 88, 2021, doi: 10.24269/jkt.v5i1.686.

PENGELOMPOKAN WILAYAH INDONESIA BERDASARKAN INDIKATOR SOSIO-EKONOMI TAHUN 2021 DENGAN MENGGUNAKAN METODE K-MEANS

Dini Fariha^{*1}, Derry Hartawan², Muhammad Al-Abrou^{*3}, Muhammad Raffi Shafwan⁴, Munirul ula⁵

¹²³⁴⁵Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹dini.210170167@mhs.unimal.ac.id, ²derry.210170161@mhs.unimal.ac.id, ³muhhammad.210170156@mhs.unimal.ac.id, ⁴muhammad.210170157@mhs.unimal.ac.id, ⁵munirulua@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk mengelompokkan wilayah Indonesia berdasarkan indikator sosio-ekonomi tahun 2021 menggunakan metode K-Means. Data yang digunakan meliputi berbagai indikator seperti jumlah penduduk, tingkat kemiskinan, tingkat pendidikan, dan pengangguran di 34 provinsi di Indonesia. Metode K-Means dipilih untuk mengidentifikasi pola-pola tersembunyi dalam data dan mengelompokkan provinsi yang memiliki karakteristik sosio-ekonomi serupa. Hasil analisis menunjukkan adanya beberapa kelompok wilayah yang berbeda berdasarkan indikator-indikator tersebut, yang dapat memberikan wawasan penting bagi pemerintah dalam perencanaan kebijakan pembangunan yang lebih tepat sasaran. Penelitian ini juga menggarisbawahi potensi pemanfaatan clustering dalam analisis data sosio-ekonomi pada skala regional.

Keywords: *K-Means, sosio-ekonomi, pengelompokan wilayah, Indonesia, kemiskinan, pendidikan, pengangguran, 2021.*

1. PENDAHULUAN

Indonesia merupakan negara kepulauan dengan keberagaman sosial dan ekonomi yang luas, mempengaruhi berbagai aspek kehidupan masyarakat di setiap wilayah. Pada tahun 2021, berbagai indikator sosio-ekonomi di Indonesia menunjukkan perbedaan yang signifikan di antara provinsi dan kabupaten/kota. Indikator ini mencakup aspek-aspek krusial seperti pendapatan per kapita, tingkat pendidikan, akses kesehatan, dan infrastruktur, yang berkontribusi pada perbedaan dalam kualitas hidup antar wilayah.

Dalam rangka memahami dan mengelola perbedaan ini, pengelompokan wilayah berdasarkan indikator sosio-ekonomi menjadi penting. Metode K-Means merupakan salah satu teknik klusterisasi yang efektif dalam mengelompokkan data berdasarkan kemiripan karakteristik. Dengan metode ini, wilayah-wilayah yang memiliki kesamaan dalam indikator sosio-ekonomi dapat dikelompokkan menjadi kluster-kluster yang homogen. Pendekatan ini tidak hanya membantu dalam memahami pola-pola sosio-ekonomi, tetapi juga dalam merancang kebijakan yang lebih terfokus dan efektif sesuai dengan kebutuhan masing-masing kluster.

Beberapa studi terkini menunjukkan bahwa penerapan metode K-Means dalam pengelompokan wilayah berdasar indikator sosio-ekonomi dapat memberikan insight yang signifikan untuk perencanaan pembangunan. Misalnya, penelitian oleh Siahaan et al. (2020) menerapkan K-Means untuk analisis kesenjangan sosial di Indonesia dan menemukan pola-pola baru dalam distribusi sumber daya. Selain itu, penelitian oleh Suhardjanto dan Widodo (2021) menggunakan teknik klusterisasi untuk mengevaluasi dampak kebijakan ekonomi di berbagai daerah. Studi-studi ini menunjukkan bahwa metode K-Means dapat mengungkap hubungan yang kompleks dalam data sosio-ekonomi dan mendukung pengambilan keputusan berbasis data.

Dalam penelitian ini, kami akan menerapkan metode K-Means untuk mengelompokkan wilayah di Indonesia pada tahun 2021 berdasarkan indikator sosio-ekonomi. Tujuan dari penelitian ini adalah untuk mengidentifikasi kluster-kluster wilayah yang memiliki karakteristik sosio-ekonomi serupa, serta untuk memberikan wawasan yang mendalam tentang distribusi kesenjangan sosial dan ekonomi di seluruh Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap strategi perencanaan dan pengembangan regional yang lebih adaptif dan responsif terhadap kebutuhan lokal.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dataset “*Socio-Economic of Indonesia in 2021*” yang tersedia di Kaggle ([tautan dataset](#)). Dataset ini mencakup beberapa indikator sosio-ekonomi yang relevan seperti:

- Provinsi
- Kabupaten/Kota
- Persentase Penduduk Miskin
- Produk Domestik Regional Bruto
- Angka Harapan Hidup
- Rata-rata Lama Sekolah
- Pengeluaran per Kapita

Dataset ini mencakup 34 Provinsi di Indonesia pada tahun 2021 dan memberikan gambaran lengkap tentang kondisi sosial dan ekonomi di setiap wilayah.

2.2 Penerapan Metode K-Means

a. Prapemrosesan Data

Prapemrosesan data merupakan langkah penting dalam menerapkan K-Means. Data yang digunakan harus dipastikan bersih dari missing values, outlier, dan dinormalisasi agar skala variabel konsisten.

- **Memuat Data:** Data sosio-ekonomi Indonesia tahun 2021 diperoleh dari dataset Kaggle yang berjudul *Socio-Economic Stats Indonesia 2021*. Dataset ini berisi data provinsi, termasuk indikator sosio-ekonomi.
- **Mengatasi Missing Values:** Jika terdapat data yang hilang, nilai tersebut diimputasi dengan menggunakan rata-rata dari indikator terkait.
- **Normalisasi Data:** Karena setiap indikator sosio-ekonomi memiliki skala yang berbeda (misalnya, GDP per capita memiliki rentang yang lebih besar dibandingkan tingkat kemiskinan), normalisasi data perlu dilakukan. Teknik yang digunakan adalah **Min-Max Scaling** untuk mengubah data menjadi rentang 0 hingga 1.

b. Menentukan Jumlah Kluster yang Optimal

Jumlah kluster yang optimal harus ditentukan sebelum K-Means dijalankan. Salah satu metode yang digunakan untuk menentukan jumlah kluster adalah **Elbow Method**. Elbow Method mengevaluasi variasi dalam data (SSE - Sum of Squared Errors) seiring bertambahnya jumlah kluster, dan mencari titik di mana penurunan SSE mulai melambat (titik elbow).

c. Pelaksanaan Klusterisasi K-Means

Setelah jumlah kluster optimal ditentukan (misalnya, jumlah optimal adalah 4 kluster), algoritma K-Means diterapkan pada data sosio-ekonomi yang telah dinormalisasi. Algoritma K-Means akan membagi wilayah Indonesia ke dalam kluster berdasarkan kemiripan dalam indikator sosio-ekonomi.

d. Interpretasi Hasil Klusterisasi

Setelah proses klusterisasi selesai, langkah berikutnya adalah menginterpretasikan hasil klusterisasi. Data yang telah dikelompokkan berdasarkan kluster dapat dianalisis lebih lanjut untuk melihat karakteristik tiap kluster. Misalnya, wilayah yang memiliki akses layanan kesehatan yang baik dan tingkat pengangguran rendah mungkin berada di satu kluster, sementara wilayah dengan tingkat kemiskinan tinggi dan akses pendidikan rendah mungkin berada di kluster lain.

e. Visualisasi Hasil Klasterisasi

Hasil klasterisasi dapat divisualisasikan untuk memudahkan interpretasi. Salah satu cara visualisasi adalah menggunakan scatter plot untuk menampilkan hubungan antar variabel berdasarkan klaster. Selain scatter plot, visualisasi geospasial juga bisa diterapkan jika tersedia koordinat geografis dari setiap provinsi, sehingga klaster dapat dilihat dalam peta wilayah Indonesia.

3. HASIL DAN PEMBAHASAN

3.1 DATA AWAL

Sebelum dilakukan proses pengelompokan, data yang digunakan dalam penelitian ini adalah **data mentah** terkait indikator-indikator sosio-ekonomi dari seluruh provinsi di Indonesia pada tahun 2021.

Tabel 1. Data Awal Sosio-Ekonomi 2021

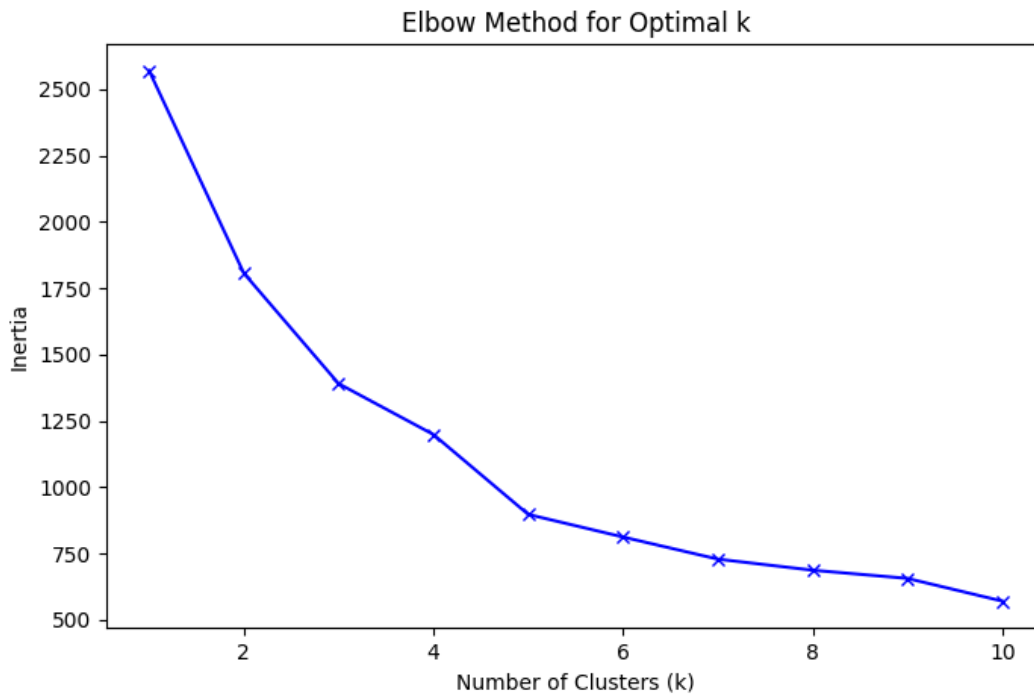
province	Cities_reg	Poor_people	Reg_gdp	Life_exp	Avg_school	Exp_percap
Aceh	simelue	18.98	2.275	65.24	9.48	7148
Aceh	Aceh Singkil	20.36	2.425	67.355	8.68	8776
Aceh	Aceh Selatan	13.18	5.531	64.36	8.88	8180
Aceh	Aceh Tengah	13.41	5.063	68.155	9.67	8030
Aceh	Aceh Timur	14.45	10.616	68.705	8.21	8577
.....
Papua	Kota Jayapura	11.39	32.207	70.455	11.57	14937

3.2 Processing Data

Tabel 1. Data Hasil Sosio-Ekonomi 2021

province	Cities_reg	Poor_people	Reg_gdp	Life_exp	Avg_school	Exp_percap	cluster
Aceh	simelue	18.98	2.275	65.24	9.48	7148	1
Aceh	Aceh Singkil	20.36	2.425	67.355	8.68	8776	1
Aceh	Aceh Selatan	13.18	5.531	64.36	8.88	8180	1
Aceh	Aceh Tengah	13.41	5.063	68.155	9.67	8030	1
Aceh	Aceh Timur	14.45	10.616	68.705	8.21	8577	1
.....
Papua	Kota Jayapura	11.39	32.207	70.455	11.57	14937	2

Adapun hasil kluster yang didapatkan setelah data diolah adalah sebanyak 3 kluster, di mana masing-masing kluster merepresentasikan karakteristik dan pola data yang berbeda-beda. Pembagian ini memungkinkan kita untuk mengelompokkan data dengan lebih efektif, sehingga setiap kluster dapat memberikan wawasan spesifik sesuai dengan ciri-ciri yang teridentifikasi. Kluster ini dapat digunakan untuk analisis lebih lanjut atau sebagai dasar dalam pengambilan keputusan strategis.



PERSAMAAN MATEMATIKA

Untuk mencari hasil perhitungan dari metode k-means, kita dapat menggunakan rumus persamaan sebagai berikut adalah rumus untuk menghitung jarak:

$$d(x_i, \mu_j) = \sqrt{\sum_{m=1}^n (x_{im} - \mu_{jm})^2} \quad (1)$$

Keterangan:

n = banyaknya atribut

x = vektor atribut real suatu data

y = vektor atribut hasil perhitungan (output) suatu data

$d(x, y)$ = jarak euclidean dari x dan y

4. DISKUSI

Hasil dari penelitian ini menunjukkan bahwa metode K-Means mampu mengelompokkan wilayah Indonesia berdasarkan indikator sosio-ekonomi pada tahun 2021 menjadi beberapa kluster yang memiliki karakteristik serupa. Penggunaan indikator sosio-ekonomi seperti **pendapatan per kapita (GDP per capita)**, **tingkat pengangguran**, **tingkat kemiskinan**, **tingkat pendidikan**, dan **akses layanan kesehatan** memberikan gambaran yang lebih komprehensif tentang kondisi sosial dan ekonomi di berbagai provinsi di Indonesia.

4.1 Perbandingan dengan Penelitian Lain

Penelitian ini sejalan dengan beberapa penelitian sebelumnya yang juga memanfaatkan metode K-Means untuk pengelompokan wilayah berdasarkan berbagai indikator sosio-ekonomi. Hasil yang diperoleh menunjukkan konsistensi dengan studi lain yang menemukan adanya disparitas yang cukup besar antara wilayah barat dan timur Indonesia, terutama dalam hal akses terhadap layanan pendidikan dan kesehatan. Namun, penelitian ini memberikan kontribusi lebih lanjut dengan menggunakan data terkini (tahun 2021) yang belum banyak digunakan dalam penelitian sebelumnya, sehingga memberikan perspektif baru dalam konteks pembangunan ekonomi pasca-pandemi COVID-19.

4.2 Keterbatasan Penelitian

Meskipun metode K-Means memberikan hasil yang cukup memuaskan, penelitian ini memiliki beberapa keterbatasan. Salah satu keterbatasan utama adalah keterbatasan data yang digunakan, yaitu data sosio-ekonomi yang tersedia hanya untuk tahun 2021. Kondisi sosio-ekonomi dapat berubah seiring waktu, sehingga pengelompokan ini bisa menjadi kurang relevan di masa mendatang. Oleh karena itu, penelitian lanjutan dengan data yang lebih aktual dan mencakup periode yang lebih panjang diperlukan untuk memantau perubahan dalam pengelompokan wilayah.

Selain itu, penggunaan indikator yang terbatas juga menjadi tantangan. Meskipun indikator yang digunakan dalam penelitian ini mencakup aspek utama dari kondisi sosio-ekonomi, ada faktor lain seperti **infrastruktur**, **keamanan**, atau **pengaruh lingkungan** yang juga bisa berperan dalam mengelompokkan wilayah. Penelitian selanjutnya dapat mempertimbangkan penambahan variabel-variabel ini untuk memperkaya hasil klusterisasi.

5. KESIMPULAN

Setelah dilakukan pengelompokan menggunakan metode K-Means, wilayah Indonesia berhasil dibagi menjadi tiga kluster berdasarkan indikator sosio-ekonomi yang relevan pada tahun 2021. Setiap kluster menunjukkan perbedaan karakteristik ekonomi dan sosial antarwilayah, yang mencerminkan variasi dalam tingkat pembangunan, kesejahteraan, serta akses terhadap layanan publik.

1. **Kluster 1:** Berisi wilayah dengan tingkat sosio-ekonomi yang relatif tinggi, seperti tingkat pendapatan per kapita yang baik, akses terhadap pendidikan dan layanan kesehatan yang memadai, serta tingkat pengangguran yang lebih rendah.
2. **Kluster 2:** Wilayah dalam kluster ini menunjukkan tingkat perkembangan sosio-ekonomi yang sedang, dengan indikator seperti pendapatan, pendidikan, dan kesehatan yang berada di tingkat menengah.
3. **Kluster 3:** Berisi wilayah yang memiliki tantangan besar dalam hal sosio-ekonomi, seperti rendahnya pendapatan, terbatasnya akses pendidikan dan layanan kesehatan, serta angka pengangguran yang relatif tinggi.

Pembagian kluster ini membantu memberikan pemahaman yang lebih jelas terkait disparitas pembangunan antarwilayah di Indonesia. Hasil kluster ini juga bisa menjadi acuan penting bagi pemerintah dan pihak terkait dalam merumuskan kebijakan yang lebih tepat sasaran guna mengurangi ketimpangan dan meningkatkan kesejahteraan masyarakat di wilayah yang tertinggal.

DAFTAR PUSTAKA

- [1] Drl, I. R., Chrisnanto, Y. H., & Umbara, F. R. (2022). ANALISIS CLUSTER PADA KELOMPOK MASYARAKAT YANG RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN METODE K-MEANS CLUSTERING DAN VISUALIASI DENGAN SIG. *Informatics and Digital Expert (INDEX)*, 4(2), 61-69.
- [2] Tiasti, R. N. E., Hamid, A., & Novitasari, D. C. R. (2021). Implementasi Algoritma K-Means Clustering untuk Pengelompokan Penyebaran Penyakit Diabetes Melitus di Kecamatan Bagor. *Jurnal Algebra*, 2(1), 69-84.
- [3] SETYANINGRUM, G. (2023). *IMPLEMENTASI METODE K-MEANS CLUSTERING DAN NAIVE BAYES CLASSIFIER DALAM MEMREDIKSI LEVEL KASUS COVID-19 DI INDONESIA* (Doctoral dissertation, UPN" Veteran" Yogyakarta).
- [4] Gustiane, I. T., Martanto, M., & Suprpti, T. (2024). CLUSTERING HASIL CEK DARAH DIABETES LANSIA MENGGUNAKAN METODE K-MEANS DI POSBINDU KP. LEBAKJERO DESA CIHERANG. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2125-2129.
- [5] Kholil, K. A., Rahaningsih, N., & Dana, R. D. (2024). PENERAPAN DATA MINING UNTUK CLUSTERING PENYAKIT DIARE MENGGUNAKAN ALGORITMA K-MEANS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3124-3131.
- [6] Rahmawati, L. (2023). Penerapan Data Mining untuk Menentukan Penyebab Kematian di Indonesia Menggunakan Metode Clustering K-Means. *PENERAPAN DATA MINING UNTUK MENENTUKAN*

PENYEBAB KEMATIAN DI INDONESIA MENGGUNAKAN METODE CLUSTERING K-MEANS, 4(3), 535-543.

- [7] Tuasikal, N. J. (2018). IMPLEMENTASI WEBGIS, GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR), DAN CLUSTERING K-MEANS UNTUK ANALISIS DATA TUBERKULOSIS DUNIA (Studi Kasus: Kasus Tuberkulosis Dunia Tahun 2015).
- [8] Ali, A., & Masyfufah, L. (2021). Klasterisasi Pasien BPJS Dengan Metode K-Means Clustering Guna Menunjang Program Jaminan Kesehatan Nasional Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *Jurnal Wiyata: Penelitian Sains dan Kesehatan*, 8(1), 8-22.
- [9] Prasatya, A., Siregar, R. R. A., & Arianto, R. (2020). Penerapan Metode K-Means Dan C4. 5 Untuk Prediksi Penderita Diabetes.

SENASTIKA Universitas Malikussaleh

PENERAPAN METODE FINITE STATE MACHINE DALAM GAME "KETEGUHAN LISA" MENGGUNAKAN RPG MAKER MV DAN ANDROID STUDIO

Ahmad Amri¹, Zara Yunizar², Hafizh Al Kautsar Aidilof³

^{1,2,3} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: ¹ ahmad765amri@gmail.com, ² zarayunizar@unimal.ac.id, ³ hafizh@unimal.ac.id

Abstrak

Perkembangan teknologi telah mendorong popularitas permainan video di kalangan anak-anak, namun banyak dari *Game* yang ada saat ini kurang memberikan nilai edukatif. Penelitian ini bertujuan untuk mengembangkan *Game* edukasi berbasis *Role-Playing Game (RPG)* yang menggunakan metode *Finite State Machine (FSM)* dalam pengembangan alur cerita dan logika permainan. *Game* "Keteguhan Lisa" dirancang menggunakan *RPG Maker MV* dan *Android Studio*, bertujuan untuk meningkatkan keterampilan kognitif dan literasi digital siswa Sekolah Dasar. Pengujian dilakukan dengan metode *Black Box Testing* dan evaluasi pengalaman pengguna melalui *System Usability Scale (SUS)*. Hasil penelitian menunjukkan bahwa *Game* ini diterima dengan baik oleh siswa, dengan skor *SUS* sebesar 78.5, yang tergolong dalam kategori "Good". *Game* ini berhasil menggabungkan elemen edukasi dan hiburan, menjadikannya media pembelajaran yang interaktif dan menyenangkan.

Keywords: *Game Edukasi; Finite State Machine; RPG; Pembelajaran Digital; Android.*

1. PENDAHULUAN

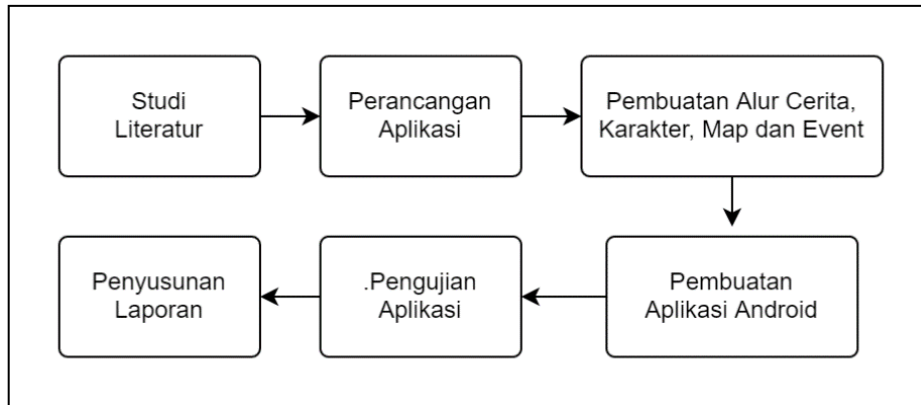
Perkembangan teknologi yang pesat telah membawa dampak signifikan terhadap berbagai aspek kehidupan, termasuk dalam hal Pendidikan [1]. Di era digital saat ini, permainan video atau *Game* menjadi salah satu bentuk hiburan yang sangat digemari, terutama oleh anak-anak [2]. Namun, di sisi lain, banyak dari *Game* tersebut lebih berfokus pada aspek hiburan semata tanpa memberikan manfaat edukatif yang signifikan [3]. Hal ini menciptakan tantangan bagi para pendidik untuk menghadirkan media pembelajaran yang tidak hanya interaktif, tetapi juga menarik minat anak-anak [4].

Salah satu genre *Game* yang populer adalah *Role-Playing Game (RPG)*, di mana pemain dapat mengendalikan karakter utama dan mengikuti alur cerita yang ditentukan oleh permainan [5]. Genre ini memiliki potensi besar untuk dimanfaatkan sebagai media edukasi karena mampu menggabungkan aspek hiburan dengan pembelajaran [6]. Metode *Finite State Machine (FSM)* digunakan dalam pengembangan *Game* untuk mengatur alur cerita dan logika transisi antar *Event* dalam *Game* [7]. *FSM* memudahkan pengembangan alur *Game* yang dinamis dan logis, di mana setiap aksi pemain memicu respons tertentu dari sistem, menciptakan pengalaman bermain yang interaktif [8].

Dalam penelitian ini, dikembangkan sebuah *Game* edukasi berbasis *RPG* yang berjudul "Keteguhan Lisa" menggunakan *RPG Maker MV* dan *Android Studio*. *Game* ini dirancang untuk meningkatkan literasi digital dan keterampilan kognitif siswa Sekolah Dasar (SD), dengan memanfaatkan elemen *puzzle* dan cerita yang menarik. Tujuan utama dari penelitian ini adalah untuk menggabungkan elemen hiburan dan edukasi dalam satu platform, serta mengukur efektivitas *Game* sebagai media pembelajaran melalui pengujian *usability* menggunakan *System Usability Scale (SUS)*.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi *Research and Development (R&D)* untuk mengembangkan *Game* edukasi berbasis *Role-Playing Game (RPG)* dengan penerapan metode *Finite State Machine (FSM)*. Tahapan penelitian meliputi perancangan, implementasi, dan pengujian *Game* "Keteguhan Lisa" yang dikembangkan menggunakan *RPG Maker MV* dan *Android Studio*. Proses pengembangan *Game* dilakukan secara bertahap dan sistematis untuk mencapai hasil yang optimal.



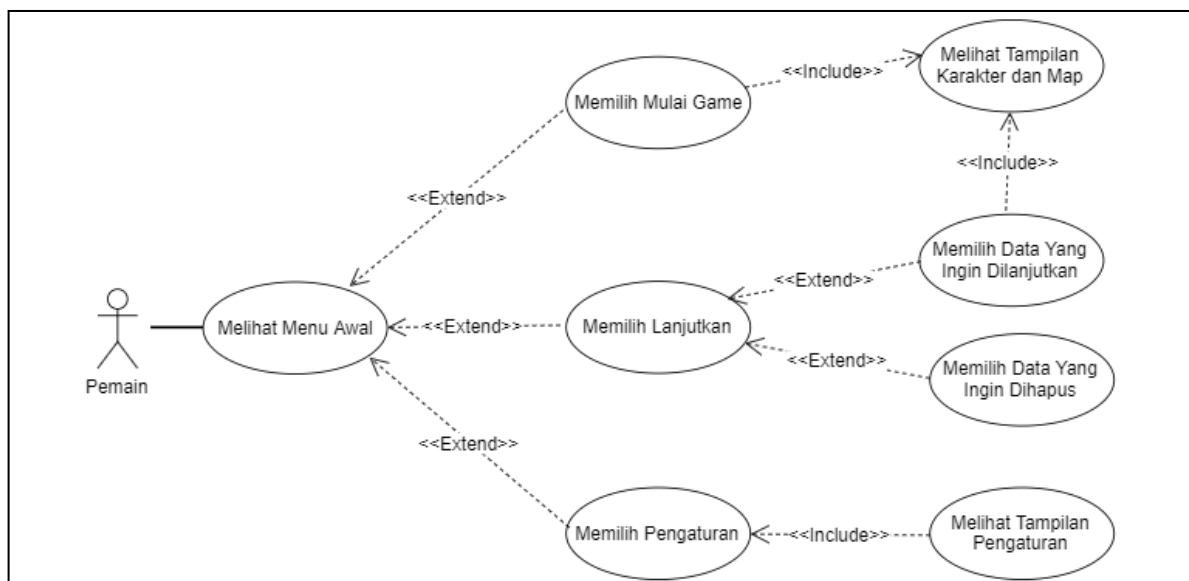
Gambar 1. Alur Penelitian

2.1. Studi Literatur

Tahap ini diisi dengan mempelajari semua kajian literatur dan teori yang ada di internet seperti jurnal, artikel, buku dan video pembelajaran yang ada.

2.2. Perancangan Aplikasi

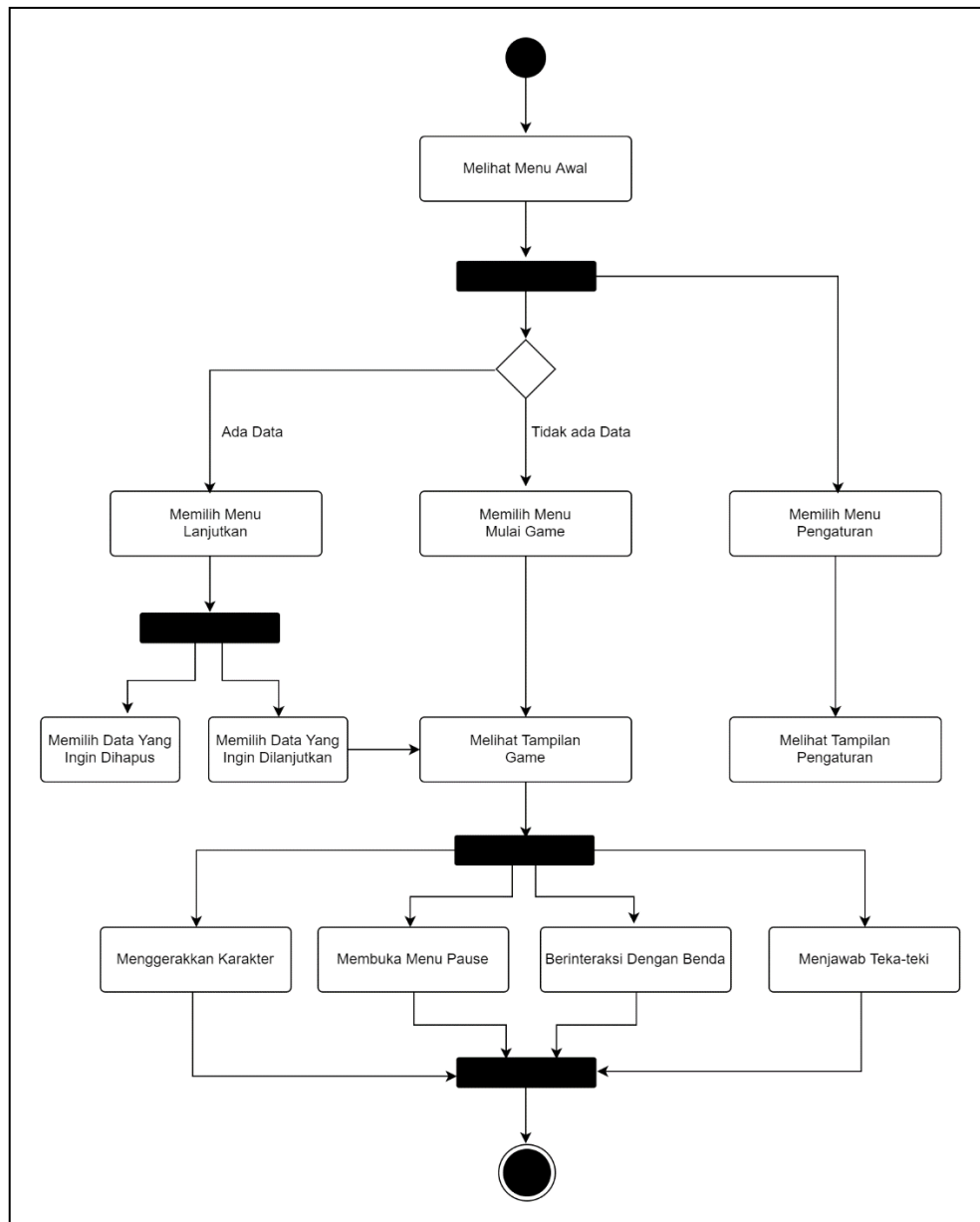
Tahap ini diisi dengan merancang sistem dan desain *Game*. Perancangan sistem *Game* akan dibuat menggunakan *UML. Unified Modelling Language* atau *UML* yang digunakan ada dua yaitu *Use Case Diagram* dan *Activity Diagram*.



Gambar 2. Use Case Diagram Menu Awal

Penjelasan usecase sebagai berikut:

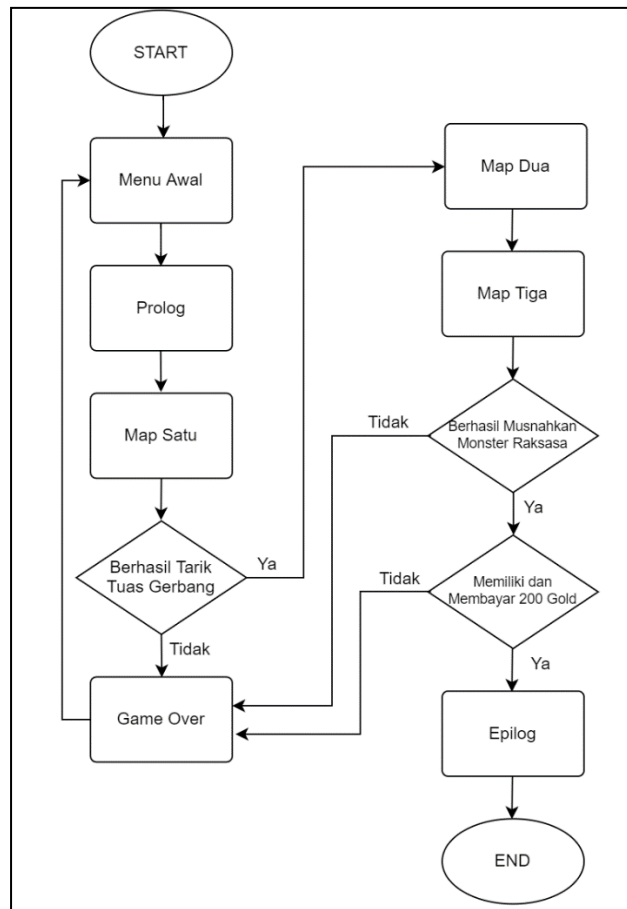
1. Pemain menekan menu *Mulai Game* akan memulai proses *Loading* sistem dan menampilkan *cut scene* prolog dari cerita *Game* *Keteguhan Lisa*.
2. Pemain menekan menu *Lanjutkan* akan memproses data *Save Game* sebelumnya dan memilih data *Save Game* yang ada untuk melanjutkan kembali permainan sebelumnya.
3. Pemain menekan menu *Pengaturan* akan menampilkan menu *options* berupa; *Comand remember*, *Synch Monitor FPS*, *BGM Volume*, *BGS Volume*, *ME Volume* dan *SE Volume*.



Gambar 3. Activity Diagram Menu Awal

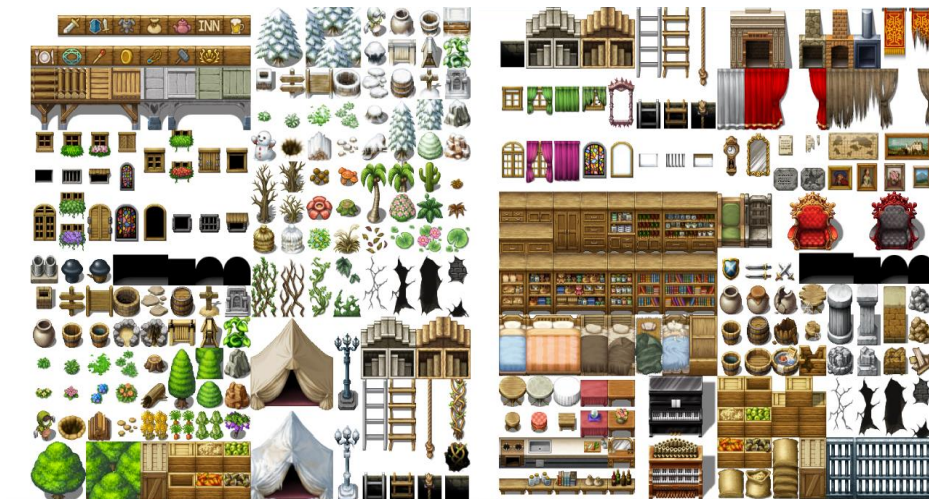
2.3. Pembuatan Alur Cerita, Karakter, Map dan Event

Tahap ini diisi dengan pembuatan semua asset *Game* seperti pembuatan *Story Game*, *character Game*, *Map Game* dan *Event Game* dengan menggunakan software *Game engine RPG Maker MV*. Tahapan ini merupakan tahapan yang membutuhkan kreatifitas yang tinggi. Untuk bisa membuat cerita yang bagus maka dibutuhkan kreatifitas yang tinggi agar para pemain tidak bosan dengan cerita yang ada serta bersemangat penuh untuk terus bermain *Game*-nya, terlebih lagi apabila cerita tersebut memiliki *plot twist* yang tidak terduga sehingga membuat pemain merasa terkejut dengan alur cerita *Game*-nya.



Gambar 4. Flowchart Rute Game

Tidak hanya membuat alur cerita yang bagus, pembuatan dari watak dan desain karakter juga akan menarik minat pemain dalam bermain *Game*. Pembuatan *Map* dari *Game* juga berpengaruh besar dalam kesenangan pemain dalam bermain *Game*. Dengan membuat *Map* yang menarik untuk dieksplorasi maka pemain tidak akan merasa bosan dengan *Gameplay RPG* yang ada.

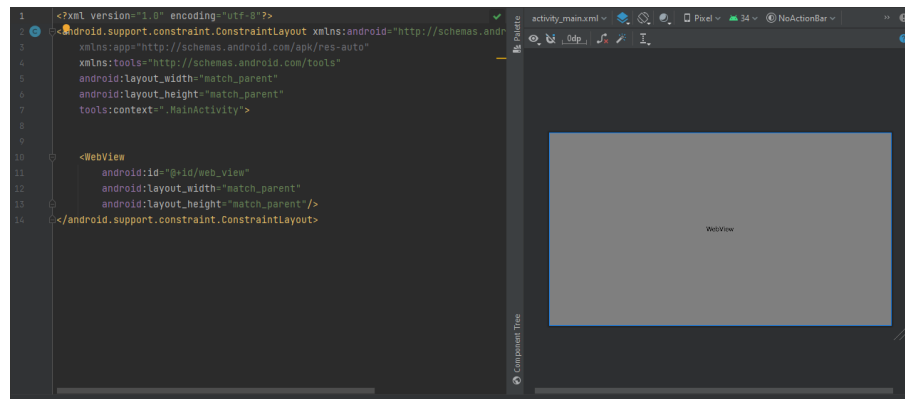


Gambar 5. Asset Map RPG Maker MV

Eksplorasi *Map* untuk mencari teka-teki dan Item merupakan salah satu dari beberapa kesenangan dalam bermain sebuah *Game RPG*. Maka dari itu dibutuhkan pembuatan *Event* yang menarik minat pemain untuk bermain seperti membuat *Event* eksplorasi *Chest* atau harta karun pada tempat tersembunyi atau disemua sudut *Map* yang ada. Pembuatan *Event* dalam pembuatan *Game* ini merupakan penerapan dari metode *Finite State Machines (FSM)* yang dimana semua *script*-nya akan terkait dengan *State*, *Event* dan *action*.

2.4. Pembuatan Aplikasi *Android*

Tahap ini diisi dengan merancang desain dan membuat program aplikasi untuk bisa menjalankan *Game* disistem operasi *Android* dengan menggunakan software *Android Studio*.



Gambar 6. Desain Di *Android Studio*

2.5. Pengujian Aplikasi

Tahap ini diisi dengan pengujian aplikasi *Game* dengan menggunakan metode *Black Box*. Setelah melakukan berbagai pengujian dengan metode *Black Box*, maka akan dilakukan *Game Tester* dengan beberapa responden. Responden akan diberi kuisioner setelah melakukan *Game Tester*, kuisioner bertujuan untuk melakukan penilaian terhadap *Game* dengan metode *SUS* atau *System Usability Scale*.

2.6. Penyusunan Laporan

Tahap ini diisi dengan penyusunan laporan dari semua perancangan *Game*, pembuatan *Game*, pengujian *Game* dan penilaian *Game* yang telah dilakukan.

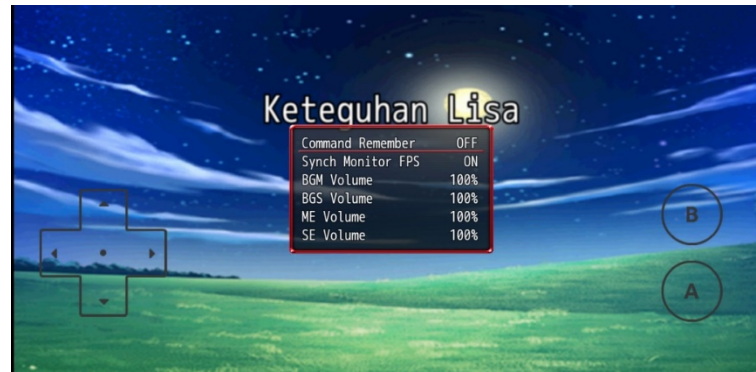
3. HASIL DAN PEMBAHASAN

3.1 Tampilan *Game*

Hasil dari pembuatan *Game* Keteguhan Lisa akan disajikan dalam beberapa tampilan dan elemen yang telah direalisasikan selama proses pengembangan *Game*. Setiap fitur yang terlibat akan ditampilkan secara visual untuk memberikan gambaran lebih jelas mengenai implementasi dari konsep yang telah dirancang sebelumnya.



Gambar 7. Tampilan Menu Awal Gambar



Gambar 8. Tampilan Menu Pengaturan

Pemain dapat memilih beberapa pengaturan dalam menu options. Pada pengaturan *Command Remember*, jika diaktifkan, sistem akan menampilkan pilihan sebelumnya saat pemain mengulangi *Event* yang sama, sedangkan jika dinonaktifkan, sistem tidak akan menunjukkan pilihan tersebut. Pada *Synch Monitor FPS*, mode *On* membuat permainan terlihat berjalan pada 60 *FPS* meskipun perangkat keras mendukung frame rate lebih tinggi, sementara mode *Off* dapat membuat permainan terlihat lambat jika perangkat keras tidak cukup kuat. Selain itu, pemain dapat mengatur volume musik latar (*BGM Volume*), suara latar belakang seperti suara burung atau air (*BGS Volume*), efek musik seperti saat memasuki dungeon (*ME Volume*), serta efek suara lain seperti ledakan atau serangan (*SE Volume*) dalam rentang 0% hingga 100%.



Gambar 9. Tampilan Menu Lanjutkan

Pemain dapat memilih file untuk menghapus atau melanjutkan progress dari permainan yang disimpan sebelumnya.



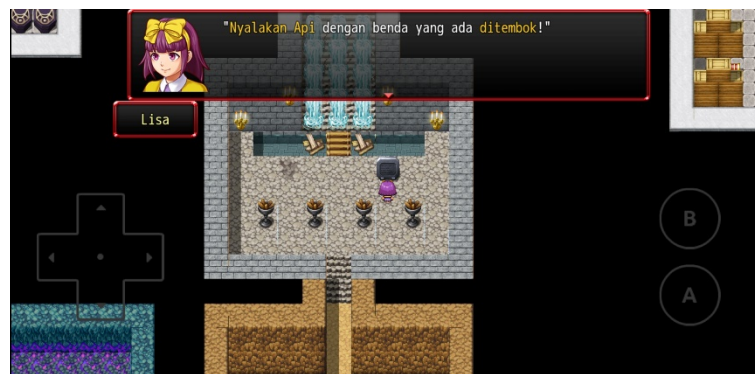
Gambar 10. Tampilan Prolog

Pemain dapat menggerakkan karakter dengan menggunakan tombol *D-pad*, berinteraksi sesuatu yang ada di *Map* menggunakan tombol *A* serta membuka menu *Pause* dengan tombol *B*.



Gambar 11. Tampilan Awal Map Satu

Pada awal permainan, muncul beberapa adegan tutorial untuk membantu pemain memahami gaya permainan dalam *Game* Keteguhan Lisa. Sebagai contoh, kotak dialog akan memberitahu pemain bahwa dengan menyentuh kristal, pemain dapat menyimpan (*Save*) data permainan. Dengan demikian, ketika pemain mengalami *Game Over*, pemain dapat memuat (*Load*) data *Game* sebelumnya.



Gambar 12. Tampilan Scene 01_03

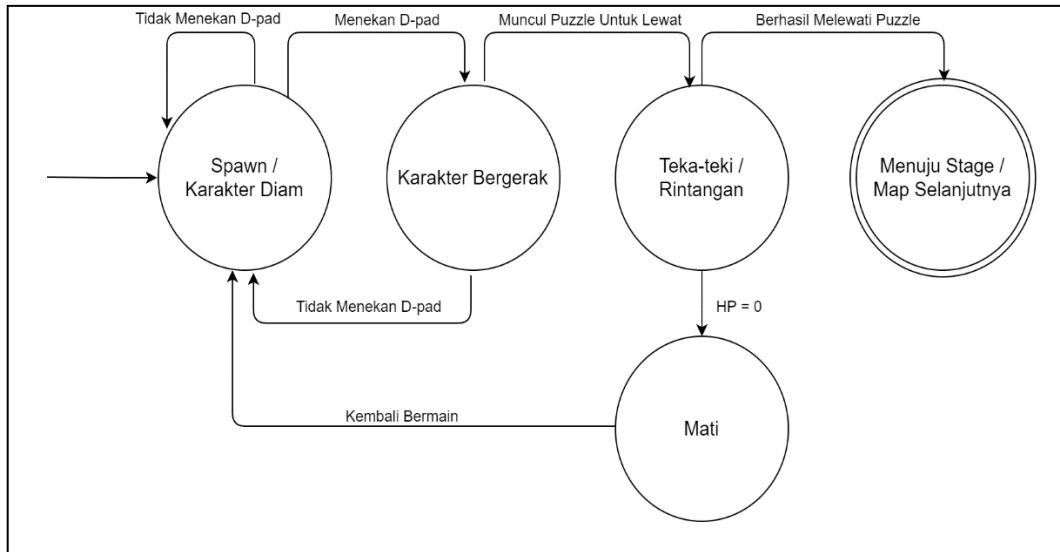
Terdapat 23 teka-teki dan rintangan didalam *Game* Keteguhan Lisa, salah satunya pada *Scene* 01_03 dimana pemain harus menghidupkan empat obor dengan menggunakan lilin yang tergantung didinding.

3.2 Implementasi *Finite State Machine*

Finite State Machine atau *FSM* adalah model mesin komputasi yang berjalan berdasarkan kondisi dan logika. Terdapat 2 logika *FSM* yang akan dipakai *Game* Keteguhan Lisa dalam software *Game engine RPG Maker MV* yaitu *Game Event Conditions* dan *Conditional Branch*. Secara konsep *Game Event Conditions* adalah sebuah kondisi dimana sebuah *script Event Game* akan aktif apabila ada salah satu dari beberapa syarat sudah terpenuhi. Sedangkan *Conditional Branch* adalah sebuah *Event* yang mempunyai kondisi logika *if else Statement* didalam *script*-nya sehingga bisa memicu beberapa *Event* yang syarat dan hasilnya bercabang.

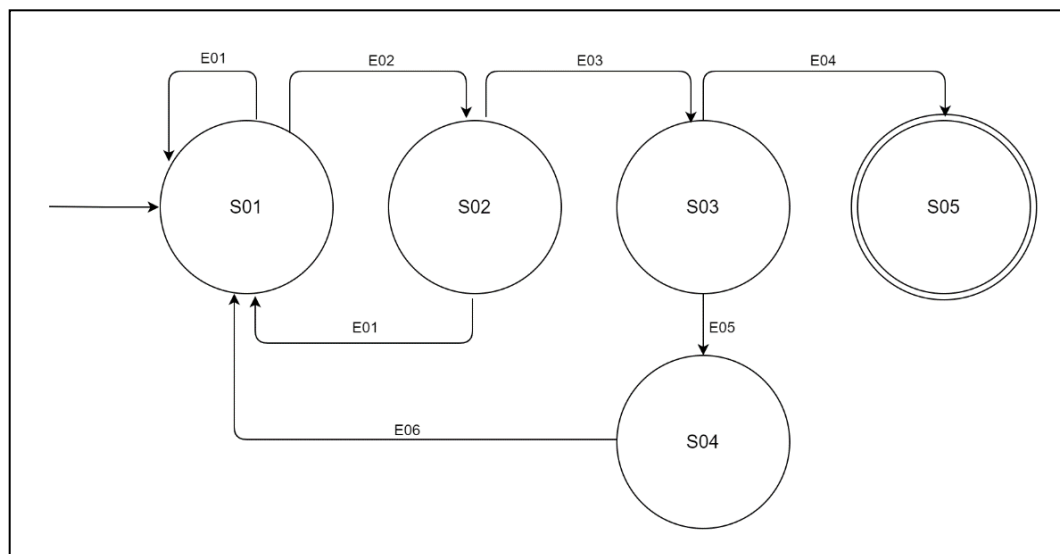
Secara umum, *script* yang menggunakan *if-else Statements* tidak secara otomatis menjadi sebuah *Finite State Machine (FSM)*. Namun, *if-else Statements* sering digunakan dalam implementasi *FSM* untuk menentukan transisi antar *State*. *FSM* biasanya melibatkan *State*, Transisi dan *Input*.

Dalam sebuah *FSM*, logika yang mengontrol transisi antar *State* dapat diimplementasikan menggunakan berbagai cara, termasuk *if-else Statements*. Perbedaannya terletak pada bagaimana struktur tersebut digunakan untuk mengelola *State* dan transisi.



Gambar 13. *FSM Game* Keteguhan Lisa

Dari Gambar 13 dapat dilihat bagaimana metode *FSM* secara keseluruhan didalam *Game* Keteguhan Lisa. Karakter yang diam akan bergerak apabila *D-pad* ditekan dan akan diam apabila *D-pad* nya dilepas. Dengan menggerakkan karakter pemain akan mengeksplorasi *Map* untuk mencari chest atau Item dan akan menjumpai sebuah teka-teki ataupun sebuah rintangan. Apabila pemain berhasil memecahkan teka-teki atau melewati rintangan, maka stage atau *Map* selanjutnya akan terbuka. Akan tetapi apabila ketika pemain tidak bisa melewati rintangan dan membuat *Health Point* karakter menjadi nol, maka karakter akan mati dan mengulang kembali ke menu awal. Jika pemain memiliki data *Save Game* maka karakter akan kembali *spawn* atau muncul di tempat *save point* berada.



Gambar 14. *DFA Game* Keteguhan Lisa

Gambar 14 menunjukkan bahwa jenis *FSM* yang dipakai *Game* Keteguhan Lisa adalah *DFA* atau *Deterministic Finite Automata*. *DFA* merupakan jenis *FSM* yang suatu *State*-nya diberi inputan maka *State* tersebut akan selalu tepat menuju satu *State*. Semua *State* dari Gambar 4.42 menunjukkan bahwa setiap jenis inputannya tidak ada yang ambigu atau setiap satu jenis inputnya hanya bertransisi ke satu *State*.

Konfigurasi dari *DFA* dapat dinyatakan sebagai berikut:

$$M = (Q, \Sigma, \delta, S, F) \quad (1)$$

Keterangan dari Persamaan 1 adalah sebagai berikut:

M : Menyatakan *FSM*.

- Q : Himpunan keadaan / *State*.
 Σ : Himpunan simbol input.
 δ : Menunjukkan fungsi transisi, memuat satu keadaan asal, satu symbol *input*, dan satu keadaan tujuan.
 S : *State* awal atau *intial State*.
 F : *Final State* atau *State* akhir.

Dengan mengikuti Persamaan 1 maka akan didapatkan konfigurasi *DFA* dari *Game* Keteguhan Lisa sebagai berikut:

- Q = {S01, S02, S03, S04, S05}
 Σ = {E01, E02, E03, E04, E05, E06}
 S = S01
 F = {S05}
 δ = $Q \cdot \Sigma$

Tabel 1. Transisi *Game* Keteguhan Lisa

Q / Σ	E01	E02	E03	E04	E05	E06
S01	S01	S02				
S02	S01		S03			
S03				S05	S04	
S04						S01
S05						

Tabel 1 merupakan tabel transisi yang dihasilkan dari rumus $Q \cdot \Sigma$, yang menggambarkan bagaimana transisi terjadi pada setiap *State* dengan berbagai input yang ada. Pada tabel ini, dapat dilihat bahwa *State* S01 akan bertransisi ke *State* S02 jika input E02 dimasukkan. Namun, jika input yang dimasukkan adalah E01, maka *State* S01 akan tetap berada pada *State* yang sama, menunjukkan bahwa tidak terjadi perubahan ke *State* lain. Hal ini berarti S01 memiliki transisi diri untuk input E01. Selain itu, dari tabel tersebut terlihat bahwa *State* S05 tidak memiliki transisi apapun dari input manapun. Tidak adanya transisi ini menandakan bahwa *State* S05 adalah *Final State*, atau *State* akhir, di mana tidak ada lagi perubahan yang dapat terjadi dari *State* tersebut dengan input apapun.

3.3 Pengujian *Black Box*

Pengujian *Black Box* dilakukan untuk memverifikasi fungsi-fungsi utama dalam *Game* Keteguhan Lisa tanpa melihat kode sumber atau struktur internal dari aplikasi. Pengujian ini bertujuan untuk memastikan bahwa setiap fitur *Game* berfungsi sesuai dengan spesifikasinya dan memberikan pengalaman yang baik kepada pengguna.

Tabel 2. Pengujian *Black Box*

No	Fungsi	Deskripsi	Expected Output	Kesimpulan
1	Buka aplikasi	Pengguna membuka aplikasi dan melihat <i>splash screen</i>	<i>Splash screen</i> muncul	Valid
2	Menu awal	Tampil menu awal (Mulai <i>Game</i> , Lanjutkan dan Pengaturan)	Menu awal tampil	Valid
3	Mulai <i>Game</i>	Pengguna memilih menu Lanjutkan	<i>Game</i> dimulai dari awal	Valid
4	Lanjutkan	Pengguna memilih menu Lanjutkan	Membuka Menu <i>Load</i>	Valid
5	Pengaturan	Pengguna memilih menu Pengaturan	Menu Pengaturan muncul	Valid
6	Tampil <i>Scene</i> 01_01	Ambil Item <i>sovel</i> untuk menghilangkan gundukan tanah	Pemain berhasil mengambil <i>sovel</i> dan menghilangkan gundukan tanah	Valid
7	Tampil <i>Scene</i> 01_02	Lewati rawa racun untuk ambil Item <i>pickaxe</i>	Pemain berhasil melewati rawa racun dan mengambil <i>pickaxe</i>	Valid
8	Tampil <i>Scene</i> 01_03	Hidupkan empat obor dengan lilin	Pemain berhasil menyalakan empat obor	Valid
9	Tampil <i>Scene</i> 01_04	Pindah ke <i>Map</i> kedua setelah menyelesaikan teka-teki di <i>Scene</i> selanjutnya	Pemain berhasil membuka gerbang dan pindah ke <i>Map</i> kedua	Valid
10	Tampil <i>Scene</i> 01_05	Turunkan semua tuas dengan warna yang sesuai	Pemain berhasil menurunkan semua tuas	Valid
11	Tampil <i>Scene</i> 01_06	Interaksi dengan buku untuk menjawab pertanyaan	Pemain menjawab pertanyaan dan mendapat poin.	Valid

Tabel 2. Pengujian *Black Box* (lanjutan)

No	Fungsi	Deskripsi	Expected Output	Kesimpulan
12	Tampil <i>Scene</i> 01_07	Hindari monster dan sentuh tugu untuk membuka pintu jeruji besi	Pemain berhasil menghindari monster dan membuka pintu jeruji besi	Valid
13	Tampil <i>Scene</i> 02_01	Nyalakan semua obor dengan kombinasi tuas yang tepat	Pemain berhasil menyalakan semua obor	Valid
14	Tampil <i>Scene</i> 02_02	Cari perbedaan antara sisi kiri dan kanan	Pemain berhasil menemukan perbedaan	Valid
15	Tampil <i>Scene</i> 02_03	Sesuaikan langkah keenam karakter agar tepat di belakang tugu	Pemain berhasil menyesuaikan langkah	Valid
16	Tampil <i>Scene</i> 02_04	Pindah ke <i>Map</i> ketiga setelah menyelesaikan teka-teki di <i>Scene</i> 02_5 dan 02_6	Pemain berhasil membuka gerbang dan pindah ke <i>Map</i> ketiga	Valid
17	Tampil <i>Scene</i> 02_05	Arahkan anak panah ke posisi yang tepat dan tarik tuasnya	Pemain berhasil mengarahkan anak panah dan menarik tuas	Valid
18	Tampil <i>Scene</i> 02_06	Melewati monster yang ada di jalan untuk menarik tuas gerbang	Pemain berhasil melewati monster dan menarik tuas	Valid
19	Tampil <i>Scene</i> 03_01	Jawab pertanyaan matematika dengan angka	Pemain berhasil menjawab pertanyaan dan menghilangkan penghalang	Valid
20	Tampil <i>Scene</i> 03_02	Jawab pertanyaan esai dengan keyboard huruf	Pemain berhasil menjawab pertanyaan esai	Valid
21	Tampil <i>Scene</i> 03_03	SUSun pipa agar terhubung dengan dua mesin	Pemain berhasil menyusun pipa dan menghubungkan dua mesin	Valid
22	Tampil <i>Scene</i> 03_04	Selesaikan <i>Mini Game</i> untuk membuka pintu	Pemain berhasil menyelesaikan <i>Mini Game</i> dan membuka pintu	Valid
23	Tampil <i>Scene</i> 03_05	Sentuh semua kristal dengan menggunakan kendaraan yang tersedia	Pemain berhasil menyentuh semua kristal	Valid
24	Tampil <i>Scene</i> 03_06	Cari empat kunci utama untuk membuka pintu	Pemain berhasil menemukan empat kunci utama dan membuka pintu	Valid
25	Tampil <i>Scene</i> 03_07	Temukan alat untuk memusnahkan monster raksasa dan kumpulkan dua kunci utama serta informasi <i>password</i> di <i>Scene</i> 03_7	Pemain berhasil menemukan alat, kunci utama, dan informasi <i>password</i> di <i>Scene</i> 03_7	Valid
26	Tampil <i>Scene</i> 03_08	Gunakan alat pemusnah monster untuk menghilangkan monster raksasa dan kumpulkan dua kunci utama di <i>Scene</i> 03_8	Pemain berhasil menggunakan alat dan mengumpulkan kunci utama di <i>Scene</i> 03_8	Valid
27	Tombol D-pad	Tombol arah untuk menggerakkan karakter atau navigasi menu	Karakter bergerak sesuai arah, atau navigasi menu berjalan lancar	Valid
28	Button A	Tombol untuk mengkonfirmasi pilihan di menu atau berinteraksi dengan objek di <i>Map</i>	Pilihan terkonfirmasi atau interaksi dengan objek terjadi	Valid
29	Button B	Tombol untuk membatalkan pilihan di menu atau membuka menu Pause di <i>Map</i>	Pilihan dibatalkan atau menu Pause terbuka	Valid
30	Tampil Prolog 01	Menampilkan awal dari cerita <i>Game</i> Keteguhan Lisa	Cerita prolog tentang Lisa dimulai	Valid

Tabel 2. Pengujian *Black Box* (lanjutan)

No	Fungsi	Deskripsi	Expected Output	Kesimpulan
31	Tampil Prolog 02	Menggambarkan Lisa masuk ke dalam lingkaran sihir teleportasi	Cerita prolog tentang Lisa masuk dungeon	Valid
32	Tampil Epilog 01	Menunjukkan kondisi ruangan setelah mencari empat kunci di <i>Scene</i> sebelumnya	Pemain melihat ruangan dengan lingkaran sihir teleportasi	Valid
33	Tampil Epilog 02	Menunjukkan Lisa kembali ke tempat semula dan terjebak di tengah lingkaran api	Cerita berakhir dengan tulisan " <i>The End?</i> "	Valid
34	Tampil Epilog 03	Menampilkan kelanjutan dari <i>Game</i> dengan menggunakan <i>Gold</i> yang diperoleh	Pemain dapat melanjutkan cerita jika memiliki cukup <i>Gold</i>	Valid
35	Tampil Epilog 04	Menampilkan Lisa pulang dengan selamat dan evaluasi diri dengan pertanyaan dari Lisa	Pemain menjawab pertanyaan dari Lisa dan melihat <i>credit Game</i>	Valid

3.4 Pengujian Responden

Untuk mengukur kegunaan *Game* "Keteguhan Lisa," dilakukan survei menggunakan *System Usability Scale* (*SUS*). *SUS* terdiri dari 10 pertanyaan dengan skala Likert 5 poin, dari 1 (sangat tidak setuju) hingga 5 (sangat setuju). Survei ini melibatkan 20 responden yang memberikan penilaian mereka terhadap berbagai aspek *Game*.

Langkah-langkah perhitungan *SUS* adalah sebagai berikut:

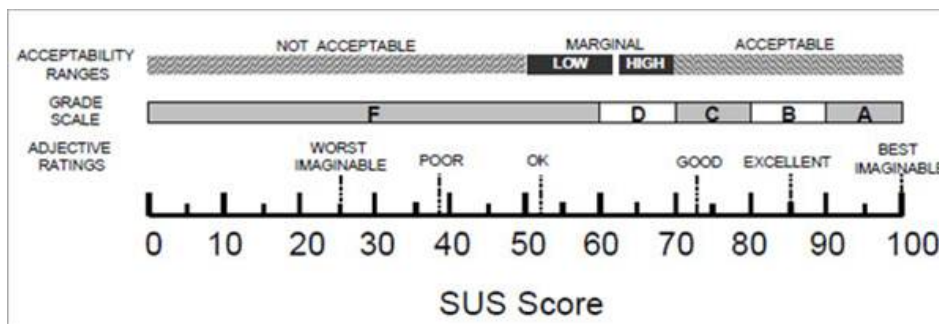
1. Setiap responden mengisi kuesioner *SUS* yang terdiri dari 10 pertanyaan.
2. Skor dari pertanyaan bernomor ganjil dikurangi 1, sedangkan skor dari pertanyaan bernomor genap dikurangi dari 5.
3. Semua nilai hasil konversi dijumlahkan untuk setiap responden.
4. Total nilai dikalikan dengan 2,5 untuk mendapatkan skor *SUS*.
5. Skor *SUS* berkisar antara 0 hingga 100, dengan skor di atas 68 dianggap dapat digunakan secara umum.

Hasil skor yang didapatkan setelah melakukan kuesioner dikelas 4 SD Negeri 5 Muara Satu, Lhokseumawe, Aceh adalah sebagai berikut.

Tabel 3. Hasil Rekapitulasi Kuesioner dan Skor

No	Responden	Skor										Jumlah	Skor (*2,5)
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
1	Responden 1	5	1	4	1	3	1	3	3	3	1	31	77,5
2	Responden 2	5	2	3	1	4	1	4	2	4	2	32	80
3	Responden 3	5	2	3	1	3	1	4	3	4	1	31	77,5
4	Responden 4	4	3	4	1	5	2	4	4	3	1	29	72,5
5	Responden 5	4	2	5	2	5	1	3	1	5	3	33	82,5
6	Responden 6	5	2	3	1	4	1	5	2	4	3	32	80
7	Responden 7	5	1	5	1	5	1	5	2	4	1	38	95
8	Responden 8	5	2	5	2	5	2	3	1	4	2	33	82,5
9	Responden 9	4	1	5	1	5	2	4	1	4	2	35	87,5
10	Responden 10	3	4	2	3	4	1	3	3	2	1	22	55
11	Responden 11	5	1	5	1	5	1	3	3	3	1	34	85
12	Responden 12	5	2	4	1	5	1	5	2	4	1	36	90
13	Responden 13	4	3	5	2	4	2	3	2	4	3	28	70
14	Responden 14	4	1	4	3	5	1	5	3	3	2	31	77,5
15	Responden 15	4	3	3	2	3	1	3	4	3	1	25	62,5
16	Responden 16	4	1	4	1	5	2	4	2	4	2	33	82,5
17	Responden 17	5	2	5	1	5	1	3	3	3	3	31	77,5
18	Responden 18	4	1	3	2	4	1	5	2	5	1	34	85
19	Responden 19	4	3	3	3	3	2	3	2	4	2	25	62,5
20	Responden 20	5	1	5	2	5	1	5	2	3	2	35	87,5
Skor Rata-Rata (Hasil Akhir)													78,5

Berdasarkan hasil penghitungan *System Usability Scale* (*SUS*) yang melibatkan 20 siswa sebagai responden, diperoleh skor akhirnya adalah sebesar 78,5.



Gambar 15. Skala Penilaian SUS

Dari Gambar 15, interpretasi skor berdasarkan beberapa skala penerimaan dapat dijelaskan sebagai berikut:

1. Berdasarkan skala penerimaan, skor 78,5 termasuk dalam kategori *Acceptable*, yang menunjukkan bahwa *Game* Keteguhan Lisa dapat diterima secara umum oleh pengguna. *Game* ini dianggap memenuhi standar kegunaan yang diperlukan bagi siswa untuk mendapatkan pengalaman yang bermanfaat selama bermain.
2. Berdasarkan skala nilai, skor 78,5 berada dalam kategori C. Hal ini menandakan bahwa *Game* Keteguhan Lisa telah mencapai tingkat kegunaan yang memadai, meskipun masih ada ruang untuk perbaikan dalam beberapa aspek. Namun, secara keseluruhan, *Game* sudah berfungsi dengan baik untuk audiens target.
3. Berdasarkan skala peringkat, skor 78,5 termasuk dalam kategori *Good*. Ini menunjukkan bahwa para pengguna, dalam hal ini siswa kelas 4 SD, menilai *Game* ini sebagai *Game* yang bermanfaat, menarik, dan mudah digunakan untuk tujuan edukatif dan hiburan. bagian ini dapat diuraikan mengenai hasil dari penelitian beserta pengujian yang telah dilakukan. Selain itu, disampaikan juga mengenai pembahasan dari penelitian maupun pengujian yang telah dilakukan.

3.4 Diskusi

Berdasarkan hasil penghitungan *System Usability Scale (SUS)* yang melibatkan 20 siswa kelas 4 SD sebagai responden, *Game* Keteguhan Lisa memperoleh skor sebesar 78,5. Skor ini termasuk dalam kategori *Acceptable* dengan peringkat *Good*, menunjukkan bahwa *Game* ini diterima dengan baik oleh pengguna dan memenuhi standar kegunaan yang dibutuhkan untuk memberikan pengalaman edukatif yang bermanfaat. Pengguna menilai *Game* ini sebagai media yang mudah digunakan dan menarik.

Ketika dibandingkan dengan penelitian lain, seperti [9] yang mengembangkan *Game* edukasi matematika berbasis *RPG Maker*, hasilnya menunjukkan bahwa *Game* ini berhasil meningkatkan pemahaman siswa sebesar 21,62%. Hal ini relevan dengan tujuan Keteguhan Lisa, yang juga berfokus pada pengembangan keterampilan siswa melalui mekanisme permainan yang melibatkan tantangan logis dan pengambilan keputusan. Meskipun Keteguhan Lisa tidak secara langsung mengukur peningkatan akademis, tingkat kegunaan yang tinggi menunjukkan potensi untuk mendukung peningkatan pemahaman siswa.

Penelitian lain, seperti [10] dan [11], juga menunjukkan keberhasilan dalam pengembangan *Game* berbasis *RPG Maker*, terutama dalam hal pengujian aset dan kegunaan. Meskipun skor *SUS* Keteguhan Lisa lebih rendah dibandingkan dengan validitas tinggi yang dicapai dalam penelitian [11] dengan skor di atas 90%, *Game* ini tetap masuk dalam kategori *Good*, menandakan keberhasilan dalam memenuhi tujuan pengembangan untuk audiens target.

Penerapan metode *Finite State Machine (FSM)* pada Keteguhan Lisa juga berhasil dalam menciptakan alur interaksi yang dinamis, mirip dengan penelitian [12], yang menemukan bahwa FSM meningkatkan perilaku cerdas NPC dalam *Game Military Defence 2D*. Penerapan *FSM* dalam Keteguhan Lisa membantu mengatur alur cerita dan interaksi yang memperkaya pengalaman bermain dan pembelajaran. Penelitian lain, seperti oleh [13], [14] serta [15], juga menekankan pentingnya *FSM* dalam menciptakan pengalaman *Game* yang logis dan interaktif, sebuah elemen yang juga berhasil diterapkan dalam Keteguhan Lisa.

Secara keseluruhan, *Keteguhan Lisa* berhasil mencapai hasil yang sejalan dengan penelitian sejenis, baik dalam aspek teknis maupun dalam dampak edukatifnya. Skor *SUS* yang cukup tinggi menunjukkan bahwa *Game* ini berhasil memenuhi tujuan pembelajaran, meskipun masih terdapat ruang untuk perbaikan pada beberapa aspek desain dan fungsionalitas. Penerapan *FSM* yang baik dalam *Game* ini juga menjadi salah satu faktor utama keberhasilan interaksi dan alur cerita yang memadai, sesuai dengan tren dalam penelitian *Game* edukasi berbasis *RPG*.

4. KESIMPULAN

Penerapan metode *Finite State Machine (FSM)* dalam pengembangan *Game* Keteguhan Lisa telah memberikan kontribusi yang signifikan dalam pengelolaan alur permainan serta interaksi pemain dengan karakter dan objek di dalam *Game*. Penggunaan *FSM* memungkinkan transisi yang terstruktur antar status permainan,

sehingga pengalaman bermain menjadi lebih dinamis dan responsif. Berdasarkan hasil pengujian yang dilakukan, baik melalui pengujian *Black Box*, maupun survei terhadap responden menggunakan *System Usability Scale (SUS)*, dapat disimpulkan bahwa *Game* Keteguhan Lisa memiliki kualitas yang baik dan dapat diterima dengan baik oleh pengguna. Pengujian *Black Box* memastikan bahwa setiap fungsi *Game* berjalan sesuai dengan spesifikasinya. Pengujian pada berbagai perangkat juga menunjukkan bahwa *Game* ini berjalan lancar tanpa masalah, baik pada perangkat dengan spesifikasi tinggi maupun menengah.

Hasil survei *SUS* yang menunjukkan skor 78,5 mengindikasikan bahwa pengguna merasa nyaman menggunakan *Game* ini, baik dari segi fungsi maupun dari segi penyampaian materi edukatif. Selain itu, responden juga merasa bahwa *Game* ini efektif dalam menggabungkan pembelajaran dengan hiburan, serta mampu meningkatkan pengetahuan dan pemahaman mereka terhadap materi yang diajarkan. Oleh karena itu, dapat disimpulkan bahwa *Game* Keteguhan Lisa berhasil memenuhi tujuan pembelajaran yang ditetapkan, serta memberikan pengalaman bermain yang memuaskan bagi para pengguna.

5. DAFTAR PUSTAKA

- [1] D. Abdullah, R. Baihaqi, Dan C. I. Erliana, “*Game* Edukasi Berbasis *Role Playing Game* Dengan Metode *Finite State Machine*,” 2015.
- [2] Antonius Fernando, G. Guntoro*, Loneli Costaner, M. Devega, Dan L. Lisnawita, “Penerapan Metode *Finite State Machine* Pada *Game* Pembelajaran Matematika,” *Jurkim*, Vol. 3, No. 1, Hlm. 60–68, Jan 2023, Doi: 10.31849/Jurkim.V3i1.11762.
- [3] A. Agustin, A. Evel, S. *SUS*Santi, Dan R. Rahmaddeni, “Implementasi Metode *Finite State Machine* Pada Permainan Tradisional Setatak Berbasis *Android*,” *Jatisi*, Vol. 8, No. 2, Hlm. 738–751, Jun 2021, Doi: 10.35957/Jatisi.V8i2.580.
- [4] M. Y. Altoofa, T. Fatimah, D. Kusumaningsih, Dan W. Pramusinto, “Implementasi *Finite State Machine* Pada *Game* ‘Malin Kundang: Simple Platform *Game*’ Dengan *Unity Game Engine*,” Vol. 2, 2023.
- [5] I. Isral, R. Tullah, N. Nurmaesah, Dan I. S. Purwanti, “Pengembangan Media Pembelajaran Seni Budaya Berbasis *Role Playing Game (RPG)* Dalam Meningkatkan Mutu Pembelajaran Siswa Sekolah Dasar,” *Ajcsr*, Vol. 5, No. 1, Hlm. 11, Jan 2023, Doi: 10.38101/Ajcsr.V5i1.607.
- [6] A. N. Gumilang, H. Haryanto, Dan E. Dolphina, “Pengembangan *Game RPG* Dan *Story* Dengan Elemen *Gameplay* Menggunakan Metodologi *Finite State Machine (FSM)* Pada *Game* Kisah Tjepoe,” *Tcv*, Vol. 1, No. 2, Hlm. 129, Jan 2024, Doi: 10.62411/Tcv.V1i2.2069.
- [7] T. Shobri, R. Akbar Hashemi W., dan S. Anwar, “Kegunaan *Game* Dalam Pengenalan Sejarah,” *Berajah Journal*, Vol. 4, 2024, doi: 10.47353/bj.v4i1.296
- [8] F. Marzian Dan M. Qamal, “*Game RPG ‘The Royal Sword’* Berbasis Desktop Dengan Menggunakan Metode *Finite State Machine (FSM)*,” *Sisfo*, Vol. 1, No. 2, Nov 2017, Doi: 10.29103/Sisfo.V1i2.244.
- [9] M. R. Rinaldi, R. Napianto, Dan M. G. An’ars, “*Game* Edukasi Berhitung Anak Sekolah Dasar Menggunakan *RPG Maker* Berbasis Mobile,” *Jurnal Teknologi Dan Sistem Informasi*, Vol. 4, No. 1, 2023.
- [10] J. M. Simanjuntak, R. Isnanto, Dan A. S. Nugroho, “Perancangan Aset Permainan ‘Satria Piningit’ Dengan Menggunakan *RPG Maker MV* Berbasis *Android*,” Vol. 2, No. 2, 2023.
- [11] S. W. Febriani, S. Sandie, Dan Y. Darma, “*Game* Edukasi Matematika Berbantuan *RPG Maker MV* Materi Bangun Datar Terhadap Kemampuan Pemecahan Masalah Matematis,” *Teorema*, Vol. 8, No. 1, Hlm. 172, Mar 2023, Doi: 10.25157/Teorema.V8i1.9936.
- [12] M. Hasan Syu’aibi, A. Mahmudi, Dan K. Auliasari, “Perancangan Dan Implementasi Metode *FSM (Finite State Machine)* Pada *Game* Military Defence 2d Berbasis *Android*,” *Jati*, Vol. 7, No. 4, Hlm. 2349–2357, Des 2023, Doi: 10.36040/Jati.V7i4.7508.
- [13] F. Mandita Dan B. Kusumo Jati, “*Application Of Finite State Machine In The 3D Game ‘Virus Hunter’*,” *Journal Of Computer Science And Visual Communication Design*, Vol. 7, No. 2, Hlm. 90–101, Des 2022, Doi: 10.55732/Jikdiskomvis.V7i2.613.
- [14] A. D. Andika, S. Mulyono, Dan B. S. W. Poetro, “*Game* Edukasi Petualangan Menggunakan *RPG Maker MV* Dengan *Finite State Machine*,” Vol. 5, No. 3, 2023.
- [15] N. F. R. Widiadhani Dan A. Solichin, “Penerapan Metode *Finite State Machine* Untuk Pergerakan Musuh Pada Permainan Edukasi ‘Petualangan Timun Mas,’” 2022.

SENASTIKA Universitas Malikussaleh

SISTEM PAKAR DIAGNOSIS PERMASALAHAN PRODUKSI TEMPE BERBASIS CASE BASED REASONING

Rayhan Alfarizi*¹, Zara Yunizar², Hafizh Al Kausar Aidilof*³

^{1,2,3} Program Studi Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: areihan56@gmail.com , zarayunizar@unimal.ac.id , hafizh@unimal.ac.id

Abstrak

Penelitian ini bertujuan mengembangkan sistem pakar berbasis *Case Based Reasoning* (CBR) untuk mendiagnosis permasalahan dalam produksi tempe, produk fermentasi penting dalam diet masyarakat Indonesia. Proses produksi tempe sering mengalami kendala seperti kegagalan fermentasi, kontaminasi mikroba, dan masalah kualitas bahan baku. Metode penelitian deskriptif ini menggunakan pendekatan kualitatif dan kuantitatif, dengan pengumpulan data melalui wawancara, observasi, studi literatur, dan kuesioner dari produsen tempe di daerah penghasil tempe. Hasil penelitian menunjukkan bahwa 60% produsen mengalami kegagalan fermentasi, 30% menghadapi kontaminasi, dan 10% mengalami masalah terkait kualitas kedelai. Sistem pakar yang dikembangkan dapat mendiagnosis masalah dan memberikan solusi dengan tingkat akurasi 85%. Penerapan sistem ini tidak hanya mempercepat proses diagnosis tetapi juga mengurangi kerugian akibat produk gagal dan meningkatkan kualitas tempe yang dihasilkan. Hasil penelitian ini menunjukkan potensi besar teknologi informasi dalam mendukung industri pangan, khususnya dalam meningkatkan efisiensi dan daya saing produk tempe di pasar. Dengan demikian, sistem pakar berbasis CBR dapat menjadi alat yang efektif bagi produsen tempe untuk mengatasi permasalahan yang dihadapi dan menjaga keberlanjutan produksi.

Keywords: *Sistem Pakar, Tempe, Fermentasi, Case Based Reasoning*

Abstract

This study aims to develop an expert system based on Case-Based Reasoning (CBR) for diagnosing issues in tempe production, an important fermented product in the Indonesian diet. The tempe production process often encounters challenges such as fermentation failure, microbial contamination, and raw material quality issues. This descriptive research employs both qualitative and quantitative approaches, with data collection conducted through interviews, observations, literature reviews, and questionnaires from tempe producers in tempe-producing regions. The research findings indicate that 60% of producers experience fermentation failure, 30% face contamination issues, and 10% encounter problems related to soybean quality. The developed expert system can diagnose problems and provide solutions with an accuracy rate of 85%. The implementation of this system not only accelerates the diagnostic process but also reduces losses due to failed products and improves the quality of the tempe produced. The results of this research demonstrate the significant potential of information technology in supporting the food industry, particularly in enhancing the efficiency and competitiveness of tempe products in the market. Thus, the CBR-based expert system can serve as an effective tool for tempe producers to address the challenges they face and ensure sustainable production.

Keywords: *Expert System, Tempe, Fermentation, Case-Based Reasoning*

1. PENDAHULUAN

Tempe adalah salah satu produk fermentasi kedelai yang kaya akan protein dan nutrisi. Di Indonesia, tempe bukan hanya menjadi sumber protein yang terjangkau, tetapi juga merupakan bagian integral dari budaya dan tradisi kuliner. Proses produksi tempe yang sederhana memungkinkan masyarakat, terutama petani dan pengusaha kecil, untuk memproduksinya secara mandiri. Namun, seperti halnya proses produksi lainnya, produksi tempe juga menghadapi berbagai permasalahan, mulai dari bahan baku, teknik fermentasi, hingga kondisi penyimpanan [1].

Dalam beberapa tahun terakhir, permintaan akan tempe meningkat, seiring dengan kesadaran masyarakat akan pentingnya konsumsi makanan sehat dan bergizi. Namun, peningkatan permintaan ini tidak selalu diimbangi dengan peningkatan kualitas dan kuantitas produksi. Banyak produsen tempe menghadapi masalah dalam proses produksi yang dapat mengakibatkan penurunan kualitas tempe, seperti perubahan rasa, bau, dan tekstur. Hal ini

sering kali disebabkan oleh kurangnya pengetahuan teknis, ketidakpahaman tentang proses fermentasi, serta sulitnya mengidentifikasi masalah yang muncul [2].

Dengan perkembangan teknologi informasi, sistem pakar dapat dimanfaatkan untuk membantu produsen dalam mendiagnosis permasalahan yang muncul dalam proses produksi tempe. Sistem pakar adalah suatu sistem komputer yang dapat menyimpan pengetahuan dari seorang ahli dan menerapkannya dalam memecahkan masalah tertentu. Salah satu pendekatan yang dapat digunakan dalam pengembangan sistem pakar adalah Case-Based Reasoning (CBR), yang memungkinkan sistem untuk belajar dari pengalaman masa lalu dan menerapkannya dalam situasi baru. Melalui pendekatan ini, diharapkan dapat memberikan solusi yang tepat dan efektif bagi produsen tempe dalam mengatasi masalah yang dihadapi [3].

1.1. Tempe

Tempe adalah makanan fermentasi tradisional yang berasal dari kedelai. Proses fermentasi melibatkan jamur *Rhizopus oligosporus*, yang mengubah kedelai menjadi produk yang lebih mudah dicerna dan kaya nutrisi. Tempe kaya akan protein, serat, vitamin, dan mineral, menjadikannya pilihan makanan sehat. Proses produksi tempe melibatkan langkah-langkah yang sederhana tetapi memerlukan perhatian pada detail untuk memastikan kualitas produk akhir. Langkah-langkah dalam produksi tempe meliputi [4]:

1. Persiapan Bahan Baku: Kedelai direndam selama beberapa jam untuk memudahkan proses perebusan dan penghilangan kulit.
2. Perebusan: Setelah direndam, kedelai direbus hingga empuk, kemudian dikeringkan.
3. Inokulasi: Kedelai yang telah direbus dicampur dengan kultur jamur *Rhizopus oligosporus*, yang merupakan mikroorganisme utama dalam proses fermentasi tempe.
4. Fermentasi: Kedelai yang diinokulasi diletakkan dalam wadah yang memungkinkan pertumbuhan jamur pada suhu yang sesuai (30-35°C) selama 24-48 jam.
5. Penyimpanan: Tempe yang telah difermentasi dapat disimpan untuk distribusi atau dikonsumsi.

1.2. Sistem Pakar

Sistem pakar adalah suatu sistem komputer yang dirancang untuk meniru dan menerapkan pengetahuan serta keterampilan seorang ahli dalam menyelesaikan masalah tertentu. Sistem ini dapat digunakan untuk mendiagnosis, merencanakan, dan memberikan rekomendasi, sehingga dapat membantu pengguna dalam pengambilan keputusan. Dengan menggunakan pendekatan berbasis pengetahuan, sistem pakar berfungsi sebagai sumber informasi yang dapat diandalkan dalam situasi di mana pengetahuan manusia tidak tersedia atau sulit diakses [5].

Sistem pakar pertama kali diperkenalkan pada awal tahun 1970-an dan telah berkembang pesat sejak saat itu, dengan aplikasi di berbagai bidang seperti kesehatan, teknik, pertanian, dan bisnis. Keunggulan utama sistem pakar adalah kemampuannya untuk memberikan solusi yang cepat, akurat, dan konsisten, serta dapat beroperasi dalam lingkungan yang tidak memerlukan kehadiran manusia secara langsung [6].

Sistem pakar terdiri dari beberapa komponen utama yang saling berinteraksi untuk mencapai tujuannya:

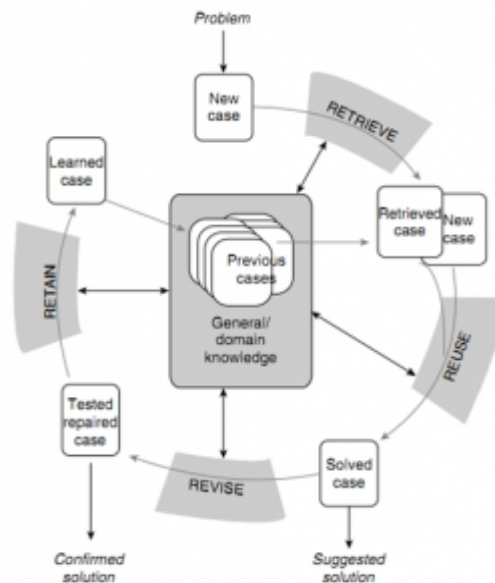
1. Basis Pengetahuan
Basis pengetahuan adalah inti dari sistem pakar. Ini merupakan kumpulan informasi, fakta, aturan, dan heuristik yang diperoleh dari para ahli di bidang tertentu.
2. Mesin Inferensi
Mesin inferensi adalah komponen yang bertanggung jawab untuk menarik kesimpulan dari informasi yang terdapat dalam basis pengetahuan. Mesin ini menggunakan algoritma tertentu untuk menganalisis data yang dimasukkan oleh pengguna dan mencocokkannya dengan informasi dalam basis pengetahuan.
3. Antarmuka Pengguna
Antarmuka pengguna adalah bagian yang berinteraksi langsung dengan pengguna. Ini memungkinkan pengguna untuk memasukkan data, mendapatkan hasil, dan berinteraksi dengan sistem pakar.
4. Basis Data
Basis data menyimpan informasi tambahan yang mungkin diperlukan oleh sistem pakar. Ini bisa termasuk data historis, statistik, atau informasi lain yang dapat membantu dalam proses pengambilan keputusan.
5. Penalaran dan Algoritma
Bagian ini mengacu pada metode atau teknik yang digunakan oleh mesin inferensi untuk memproses data dan menarik kesimpulan.

1.3. Case Based Reasoning

Case Based Reasoning (CBR) adalah pendekatan dalam kecerdasan buatan yang berfokus pada penyelesaian masalah dengan memanfaatkan pengalaman masa lalu. Dalam metode ini, sistem akan mencari solusi untuk masalah baru dengan merujuk pada kasus-kasus yang telah terjadi sebelumnya dan menyimpan solusi dari kasus-kasus tersebut. CBR didasarkan pada asumsi bahwa masalah yang dihadapi saat ini memiliki kesamaan

dengan masalah yang pernah dihadapi sebelumnya, sehingga solusi yang diterapkan pada kasus lama dapat diterapkan pada kasus baru [7].

Proses *Case Based Reasoning* (CBR) umumnya terdiri dari empat langkah utama. Berikut adalah penjelasan untuk setiap langkah:



Gambar 1. Tahapan *Case Based Reasoning*

Dalam tahap pengambilan kembali, akan dilakukan analisis untuk menghitung tingkat kesamaan struktur antara kasus baru dan objek-objek yang terdapat dalam kasus-kasus yang sudah tersimpan. Fungsi kemiripan (Similarity) digunakan sebagai metode untuk mengidentifikasi kesamaan antara kasus yang ada dalam basis data dan kasus yang baru. Rumus yang digunakan untuk mencari nilai kemiripan yaitu :

$$\text{Similarity} = \frac{s_1 * w_1 + s_2 * w_2 + ..s_n * w_n}{w_1 + w_2 + ...w_n} \dots\dots\dots(1)$$

Keterangan :

S = similarity jika terdapat kemiripan kasus maka akan bernilai 1, sedangkan tidak mirip, bernilai 0.
 W = weight (bobot yang diberikan).

2. METODE PENELITIAN

2.1. Tempat dan Waktu Penelitian

Isi dari metode penelitian adalah memformulasikan permasalahan yang diteliti dengan lebih rinci (sedapat mungkin ditulis secara matematis) dan menjelaskan metode yang diusulkan. Apabila menggunakan sebuah algoritma, dapat dijelaskan di bagian ini, beserta dengan *state of the art*.

Isi Metode Penelitian umumnya hanya mencakup 20-30% dari keseluruhan paper.

2.2. Jenis Penelitian

Penelitian ini menggunakan metode deskriptif dengan pendekatan kualitatif dan kuantitatif. Penelitian ini bertujuan untuk mengembangkan sistem pakar berbasis CBR untuk mendiagnosis masalah yang sering dihadapi dalam produksi tempe.

2.3. Tahapan Penelitian

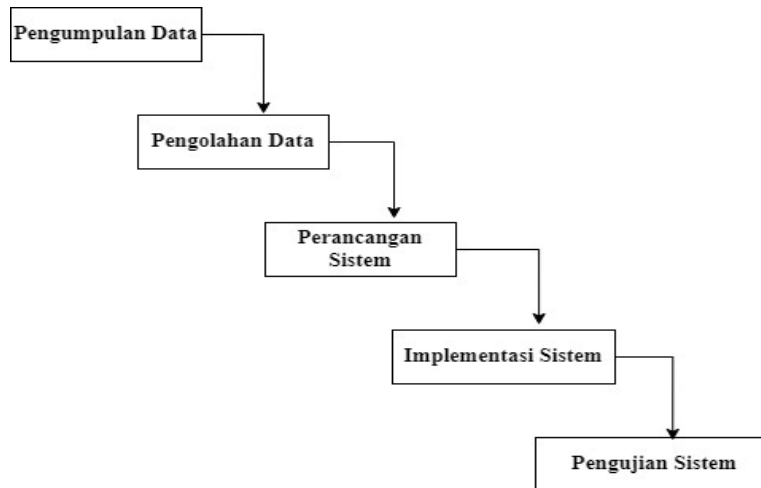
Data dikumpulkan melalui beberapa teknik:

1. Wawancara: Melakukan wawancara mendalam dengan produsen tempe untuk mengidentifikasi permasalahan yang sering mereka hadapi serta solusi yang mereka terapkan.
2. Studi Literatur: Mengkaji literatur terkait proses produksi tempe, permasalahan yang umum terjadi, dan penerapan sistem pakar dalam industri pangan.

3. Observasi: Melakukan observasi langsung terhadap proses produksi tempe untuk memahami secara detail tahapan dan permasalahan yang mungkin timbul.

2.4. Skema Penelitian

Berikut adalah skema penelitian yang akan dilakukan oleh penulis untuk memperoleh data atau informasi yang dibutuhkan. Berikut skema penelitian yang dilakukan adalah :



Gambar 2. Skema Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Sistem pakar yang dikembangkan mampu mendiagnosis permasalahan dalam produksi tempe dengan mengandalkan basis data kasus yang ada. Pengguna dapat memasukkan gejala yang dihadapi, dan sistem akan mencocokkannya dengan kasus yang pernah terjadi. Setelah dilakukan pengujian, sistem menunjukkan akurasi yang tinggi dalam mendiagnosis permasalahan. Data menunjukkan bahwa sistem dapat memberikan rekomendasi yang tepat dalam 80-90% kasus.

3.2. Data Yang Digunakan

1. Dataset

Table 1. Dataset

No	Tanggal Produksi	Jenis Kedelai	Sumber kedelai	Waktu Pemasakan (Menit)	Suhu Fermentasi (C)	Waktu Fermentasi (Jam)	Kadar Air (%)	Kualitas Kedelai	Masalah Yang Dihadapi	Solusi yang dicoba
1	05-07-2024	Kedelai Lokal	Petani A	60	30	24	55	Baik	Tidak Berfermentasi	Meningkatkan Suhu fermentasi
2	12-07-2024	Kedelai Impor	Distributor B	45	32	20	50	Sangat Baik	Bau Tidak Sedap	Mengubah Metode Penyimpanan
3	19-07-2024	Kedelai Impor	Petani C	50	28	18	53	Cukup	Tekstur Keras	Manambahkan Ragi Lebih Banyak
4	26-07-2024	Kedelai Lokal	Petani D	55	31	22	52	Baik	Pertumbuhan Jamur	Sterilisasi Alat
5	02-08-2024	Kedelai Lokal	Petani E	60	30	24	54	Baik	Tidak Berfermentasi	Mengurangi Waktu Fermentasi
6	09-08-2024	Kedelai Impor	Distributor F	40	33	21	51	Sangat Baik	Warna Merata	Menyesuaikan Kadar Air

2. Penjelasan Variabel

Keterangan :

- Tanggal Produksi = Tanggal saat tempe diproduksi.
 Jenis Kedelai = Varietas kedelai yang digunakan.
 Sumber Kedelai = Dari mana kedelai diperoleh.

Waktu Pemasakan	= Durasi pemasakan kedelai sebelum fermentasi.
Suhu Fermentasi	= Suhu ruangan atau lingkungan saat fermentasi.
Waktu Fermentasi	= Lama waktu kedelai difermentasi.
Kadar Air	= Persentase kadar air dalam tempe yang dihasilkan.
Kualitas Kedelai	= Penilaian umum tentang kualitas kedelai yang digunakan.
Masalah yang Dihadapi	= Masalah yang muncul selama proses produksi.
Solusi yang Dicoba	= Upaya yang dilakukan untuk mengatasi masalah tersebut.
Penggunaan Data	

Data di atas dapat digunakan untuk:

1. Menganalisis faktor-faktor yang mempengaruhi kualitas tempe.
2. Mengidentifikasi pola masalah yang sering muncul.
3. Mengembangkan rekomendasi untuk meningkatkan proses produksi tempe.

4.3 Perhitungan *Case Based Reasoning*

1. Analisis Rata-rata Kadar Air

Kita akan menghitung rata-rata kadar air dari semua contoh data yang ada.

Data Kadar Air

55%

50%

53%

52%

54%

51%

Perhitungan Rata-rata Kadar Air

$$\text{Rata-rata Kadar Air} = \frac{55 + 50 + 53 + 52 + 54 + 51}{6} = \frac{315}{6} = 52.5\%$$

2. Analisis Waktu Fermentasi

Menghitung rata-rata waktu fermentasi dari semua contoh data.

Data Waktu Fermentasi

24 jam

20 jam

18 jam

22 jam

24 jam

21 jam

Perhitungan Rata-rata Waktu Fermentasi

$$\text{Rata-rata Waktu Fermentasi} = \frac{24 + 20 + 18 + 22 + 24 + 21}{6} = \frac{129}{6} = 21.5 \text{ jam}$$

3. Frekuensi Masalah

Menghitung frekuensi masing-masing masalah yang muncul.

1. Data Masalah yang Dihadapi

2. Tidak Berfermentasi

3. Bau Tidak Sedap

4. Tekstur Keras

5. Pertumbuhan Jamur

6. Tidak Berfermentasi

7. Warna Tidak Merata

Frekuensi

1. Tidak Berfermentasi: 2 kali

2. Bau Tidak Sedap: 1 kali

3. Tekstur Keras: 1 kali

4. Pertumbuhan Jamur: 1 kali

5. Warna Tidak Merata: 1 kali

Analisis Hubungan Suhu dan Kualitas

Menghitung hubungan antara suhu fermentasi dan kualitas kedelai (dalam bentuk skala). Misalnya, kita dapat memberikan skor pada kualitas kedelai:

Sangat Baik = 3

Baik = 2

Cukup = 1

Data Kualitas Kedelai dan Suhu

Table 2. suhu fermentasi dan kualitas kedelai

Suhu Fermentasi (C)	Kualitas Kedelai	Skor
30	Baik	2
32	Sangat Baik	3
28	Cukup	1
31	Baik	2
30	Baik	2
33	Sangat Baik	3

Perhitungan Rata-rata Skor Kualitas

$$\text{Rata-rata Skor Kualitas} = \frac{62 + 3 + 1 + 2 + 2 + 3}{6} = \frac{613}{6} \approx 2.17$$

4. DAFTAR PUSTAKA

- [1] Achsanuddin, A., Khaerunnisa, F., Violin, V., & Yusuf, M. (2023). Faktor-faktor yang Memengaruhi Produksi Tempe di Kecamatan Burau Kabupaten Luwu Timur. *Kompeten: Jurnal Ilmiah Ekonomi dan Bisnis*, 2(2), 568-577. <https://doi.org/10.57141/kompeten.v2i2.76>
- [2] Astrawinata, Y. F., Triono, J., & Utomo, P. (2021). Sistem Pakar Penentuan Penempatan Karyawan Terhadap Bidang Pekerjaan Berbasis Web Dengan Metode Case Based Reasoning Studi Kasus CV. Mitra Teknik. *Journal of Information Technology Ampera*, 2(2), 90-104. [10.51519/journalita.volume2.isssue2.year2021.page90-104](https://doi.org/10.51519/journalita.volume2.isssue2.year2021.page90-104)
- [3] Ayustina, B., Nurdini, A., & Lazuardy, A. (2023). Perencanaan Jadwal Induk Produksi Pada Produk Tempe Di Rumah Tempe Indonesia. *Jurnal Ilmiah Teknik*, 2(1), 60-75. <https://doi.org/10.56127/juit.v2i1.497>
- [4] Falah, A. L. N., Arief, K., & Riginianto, R. S. I. (2023). Analisis Pengendalian Kualitas Pada Tempe Menggunakan Metode Seven Tools Dan FMEA. *Jurnal Teknologi dan Manajemen Industri Terapan*, 2(3), 212-223. <https://doi.org/10.55826/tmit.v2i3.264>
- [5] Fauziah, A. R. (2024). *Analisis Perhitungan Harga Pokok Produksi Tempe Dengan Metode Full Costing Pada Agroindustri Tempe Asli Hb Jambi* (Doctoral Dissertation, Universitas Unja). <https://doi.org/10.31102/agrosains.2024.9.1.31-39>
- [6] Jafar, J., Syam, A., & Utamingsih, D. (2024). Fermentasi Limbah Cair Produksi Tempe Terhadap Pertumbuhan Seledri (*Apium graveolens* L.). *Jurnal Biotek*, 12(1), 18-28. <https://doi.org/10.24252/jb.v12i1.37098>
- [7] Julyadin, M. R., & Herdiansyah, R. (2024). Sistem Pakar Untuk Mendiagnosa Tingkat Stres Pada Mahasiswa Tingkat Akhir Dengan Menggunakan Metode Case Based Reasoning (CBR) Berbasis Website. *Madani: Jurnal Ilmiah Multidisiplin*, 2(8), 562-567. <https://doi.org/10.5281/zenodo.13380136>
- [8] Kusumawati, I., Astawan, M., & Prangdimurti, E. (2020). Efisiensi Proses Produksi dan Karakteristik Tempe dari Kedelai Pecah Kulit (Production Process Efficiency and Characteristic of Tempe from Dehulled Soybean). *Jurnal Pangan*, 29(2), 117-126. <https://doi.org/10.33964/jp.v29i2.492>
- [9] Nas, C., & Kusnadi, K. (2019). Sistem Pakar Diagnosa Penyakit Gigi Dan Mulut Menggunakan Metode Case-Based Reasoning. *Jurnal Digit: Digital of Information Technology*, 9(2), 202-214. <https://doi.org/10.51920/jd.v9i2.122>
- [10] Pamungkas, G. A., Amellia, A., & Alifdiyani, A. (2024). Perencanaan Dan Pengendalian Produksi Tempe Dengan Menggunakan Metode Manufacturing Resources Planning. *Jurnal Ilmiah Penelitian Mahasiswa*, 2(4), 300-310. <https://doi.org/10.61722/jipm.v2i4.289>
- [11] Putri, D. C., & Siregar, J. (2023). Analisis Kelayakan Bisnis pada Usaha Produksi Tempe Rumahan dalam Perspektif Islam. *Jurnal Ilmu Komputer, Ekonomi dan Manajemen (JIKEM)*, 3(1), 1625-1637. <https://doi.org/10.21107/agrointek.v12i2.4025>
- [12] Rahman, S. A., & Sumijan, S. (2021). Sistem Pakar Menggunakan Metode Case Based Reasoning dalam Akurasi Penyakit Disebabkan oleh Bakteri *Staphylococcus Aureus*. *Jurnal Sistim Informasi dan Teknologi*, 13-19. <https://doi.org/10.26418/justin.v7i2.29891>
- [13] Septiany, A., & Gunawan, K. I. (2024). Analisis Kelayakan Usaha Produksi Tempe Di Desa Pandawangi Kecamatan Diwek Kabupaten Jombang. *Musyteri: Neraca Manajemen, Akuntansi, dan Ekonomi*, 4(1), 32-42. <https://doi.org/10.8734/musyteri.v4i1.2247>

- [14] Septiany, A., & Gunawan, K. I. (2024). Analisis Kelayakan Usaha Produksi Tempe Di Desa Pandawangi Kecamatan Diwek Kabupaten Jombang. *Musyteri: Neraca Manajemen, Akuntansi, dan Ekonomi*, 4(1), 32-42. <https://doi.org/10.30997/karimahtauhid.v3i10.15507>
- [15] Syahputra, M., Defit, S., & Sumijan, S. (2021). Sistem Pakar Metode Case Based Reasoning untuk Mengidentifikasi Penyakit Psoriasis. *Jurnal Sistim Informasi Dan Teknologi*, 20-27. <https://doi.org/10.37034/jsisfotek.v3i1.39>

SENASTIKA Universitas Malikussaleh

SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN OBJEK WISATA DI KABUPATEN ASAHAN MENGGUNAKAN METODE TOPSIS DAN AHP

Fauzan Habibie Dityas^{*1}, Yolinda Cesilia², Naza Amarianda³, Fakhurrrazi⁴,
Munirul Ula⁵

^{1,2,3,4} Program Studi Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹ fauzan.210170200@mhs.unimal.ac.id, ² yolinda.210170208@mhs.unimal.ac.id, ³
naza.210170206@mhs.unimal.ac.id, ⁴ fakhurrrazi.210170219@mhs.unimal.ac.id,
⁵ munirulula@unimal.ac.id

Abstrak

Pengembangan sektor pariwisata merupakan salah satu prioritas di Kabupaten Asahan, yang memiliki beragam objek wisata potensial. Namun, dalam pemilihan destinasi wisata yang sesuai, wisatawan seringkali dihadapkan pada kesulitan karena banyaknya pilihan serta perbedaan karakteristik di setiap lokasi. Oleh karena itu, diperlukan sebuah Sistem Pendukung Keputusan (SPK) yang dapat membantu proses pemilihan dengan memberikan rekomendasi yang obyektif, akurat, dan berdasarkan preferensi pengguna. Penelitian ini bertujuan untuk merancang dan mengimplementasikan SPK pemilihan objek wisata di Kabupaten Asahan dengan menggunakan metode TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) dan AHP (Analytic Hierarchy Process). Metode AHP berfungsi untuk menetapkan bobot kriteria berdasarkan tingkat kepentingan, sementara TOPSIS digunakan untuk melakukan evaluasi dan perankingan alternatif objek wisata. Kriteria evaluasi yang digunakan dalam sistem ini meliputi aksesibilitas, fasilitas, daya tarik wisata, keamanan, dan kenyamanan, yang masing-masing memiliki bobot sesuai dengan preferensi pengguna. Sistem ini tidak hanya mempermudah wisatawan dalam menentukan destinasi yang paling sesuai, tetapi juga membantu pihak terkait dalam pengelolaan dan pengembangan destinasi wisata yang lebih tepat sasaran. Hasil dari implementasi menunjukkan bahwa kombinasi metode TOPSIS dan AHP mampu menghasilkan peringkat objek wisata yang akurat dan dapat diandalkan. Dengan adanya SPK ini, diharapkan pariwisata di Kabupaten Asahan dapat berkembang lebih optimal, serta mampu menarik lebih banyak wisatawan.

Keywords: Sistem Pendukung Keputusan, TOPSIS, AHP, Destinasi Wisata, Kabupaten Asahan

Abstract

The development of the tourism sector is one of the priorities in Asahan Regency, which has a variety of potential tourist attractions. However, in selecting a suitable tourist destination, tourists are often faced with difficulties due to the many choices and differences in characteristics in each location. Therefore, a Decision Support System (DSS) is needed that can assist the selection process by providing objective, accurate recommendations based on user preferences. This study aims to design and implement a DSS for selecting tourist attractions in Asahan Regency using the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and AHP (Analytic Hierarchy Process) methods. The AHP method functions to determine the weight of criteria based on the level of importance, while TOPSIS is used to evaluate and rank alternative tourist attractions. The evaluation criteria used in this system include accessibility, facilities, tourist attractions, security, and comfort, each of which has a weight according to user preferences. This system not only makes it easier for tourists to determine the most suitable destination, but also helps related parties in the management and development of more targeted tourist destinations. The results of the implementation show that the combination of the TOPSIS and AHP methods is able to produce accurate and reliable tourist attraction rankings. With this SPK, it is hoped that tourism in Asahan Regency can develop more optimally and be able to attract more tourists.

Keywords: Decision Support System, TOPSIS, AHP, Tourist Destination, Asahan Regency

1. PENDAHULUAN

Di zaman modern ini, hidup dan bekerja di tengah tuntutan yang semakin meningkat dapat meningkatkan risiko depresi, sehingga orang membutuhkan waktu untuk menyegarkan pikiran dan mengurangi beban kerja secara menyeluruh [1]. Wisata merupakan perjalanan individu atau kelompok yang bertujuan untuk mengunjungi lokasi tertentu dengan tujuan rekreasi atau penelitian mengenai tempat tersebut, biasanya dilakukan dalam waktu yang terbatas [2]. Peningkatan sektor pariwisata harus terus ditingkatkan untuk menjadi bagian dari

ekonomi nasional dan lokal yang lebih baik [3]. Perjalanan wisata memerlukan perencanaan yang akurat untuk berpindah dari satu tempat ke tempat lainnya [4]. Sektor pariwisata memiliki peran yang sangat penting dalam mendukung pertumbuhan ekonomi suatu daerah, termasuk di Kabupaten Asahan. Dengan

2 SENASTIKA 2024, Jurusan Informatika Universitas Malikussaleh

berbagai potensi wisata yang tersebar di berbagai wilayah, Kabupaten Asahan memiliki peluang besar untuk meningkatkan jumlah kunjungan wisatawan, baik lokal maupun mancanegara. Namun, dengan banyaknya pilihan objek wisata yang tersedia, wisatawan seringkali kesulitan menentukan destinasi yang sesuai dengan preferensi dan kebutuhan mereka. Setiap objek wisata memiliki keunikan tersendiri, baik dari segi aksesibilitas, fasilitas, daya tarik, hingga aspek keamanan dan kenyamanan, sehingga pemilihan destinasi menjadi sebuah proses yang kompleks. Orang melakukan perjalanan wisata karena ingin mengenal budaya orang lain, menghindari rutinitas yang monoton, bertemu dan bersosialisasi dengan orang baru, memperkuat ikatan keluarga, dan bahkan untuk menunjukkan status social [5]. Dari berbagai motivasi perjalanan wisata tersebut, wisatawan dapat merasakan manfaat seperti mengurangi stres, membangkitkan kreativitas, meningkatkan fokus, meningkatkan kebugaran fisik dan kebahagiaan, serta memperkuat hubungan interpersonal [6]. Dengan mempertimbangkan alasan dan manfaatnya, semakin banyak orang yang melakukan perjalanan wisata setiap tahunnya, dan jumlah ini terus bertambah [1].

Untuk mengatasi hal ini, diperlukan suatu alat yang dapat memberikan rekomendasi berdasarkan analisis yang obyektif dan relevan. Sistem Pendukung Keputusan (SPK) merupakan solusi yang tepat untuk membantu proses pengambilan keputusan di berbagai bidang, termasuk pariwisata. Dengan SPK, wisatawan dapat memilih destinasi yang sesuai dengan kriteria tertentu, seperti aksesibilitas, fasilitas, daya tarik wisata, keamanan, dan kenyamanan. Selain itu, SPK juga dapat membantu pihak pengelola pariwisata dalam menetapkan kebijakan yang lebih efektif untuk pengembangan destinasi wisata.

Penelitian ini bertujuan untuk mengembangkan sistem yang mampu merekomendasikan destinasi wisata di Kabupaten Asahan secara obyektif dan efektif serta membantu dalam perumusan masalah yang dihadapi untuk menentukan destinasi wisata sehingga keputusan yang tepat dapat diambil nantinya [7]. Sistem ini diharapkan tidak hanya bermanfaat bagi wisatawan, tetapi juga bagi pemerintah daerah dalam merumuskan strategi pengembangan pariwisata yang lebih berkelanjutan dan terarah. Penelitian ini menggabungkan metode TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) dan AHP (Analytic Hierarchy Process) dalam merancang SPK pemilihan objek wisata di Kabupaten Asahan. AHP digunakan untuk menentukan bobot kriteria berdasarkan tingkat kepentingan pengguna, sementara TOPSIS digunakan untuk meranking alternatif destinasi wisata dengan membandingkannya terhadap solusi ideal. Kombinasi metode ini diharapkan dapat memberikan hasil yang lebih akurat dan mempermudah proses pengambilan keputusan.

2. METODE PENELITIAN

Penelitian ini dilakukan untuk merancang dan mengembangkan Sistem Pendukung Keputusan (SPK) yang berfungsi memberikan rekomendasi pemilihan objek wisata di Kabupaten Asahan dengan memanfaatkan metode TOPSIS dan AHP. Berikut adalah data yang diperoleh dari website <https://portal.asahankab.go.id/tempat-wisata-di-kabupaten-asahan>.

No	Nama Alternatif	Kode
1	Danau Teratai	A1
2	Air Terjun Unong Sisapa	A2
3	Air Terjun Simonang monang	A3
4	Air Terjun Ponot	A4
5	Bedeng	A5
6	Danau Dikelurahan Kisaran N	A6
7	Arung Jeram	A7
8	Pelabuhan	A8
9	Pasiran	A9

2.1. Sistem Pendukung Keputusan

Sistem Pendukung Keputusan (SPK) merupakan sistem yang menggunakan teknologi informasi dan kecerdasan buatan untuk membantu pengambilan keputusan dalam situasi yang kompleks dan ambigu [8][9]. SPK mengandalkan data yang relevan dan akurat untuk memberikan informasi yang berguna dalam pengambilan keputusan. Data ini dapat berasal dari berbagai sumber internal dan eksternal. SPK menggunakan model matematis

dan teknik analisis untuk mengolah data dan mengidentifikasi solusi atau alternatif terbaik dari berbagai pilihan yang ada [10]

2.2. Metode Analytical Hierarchy Process (AHP)

Metode ini menjabarkan masalah berupa multi kriteria yang bersifat kompleks menjadi suatu bentuk hirarki yang artinya hirarki sendiri merupakan suatu gambaran dari suatu permasalahan yang bersifat kompleks atau rumit dalam sebuah struktur multi level. Level pertama dari suatu hirarki ini adalah goals, selanjutnya level kriteria, sub kriteria, dan kebawah hingga level terakhir dari level alternatif [11]. AHP mengelompokkan kriteria penilaian dalam suatu hierarki, dan menggunakan nilai skala Saaty [12] pengganti persepsi manusia untuk melakukan perbandingan tingkat kepentingan satu kriteria dengan kriteria lainnya [13].

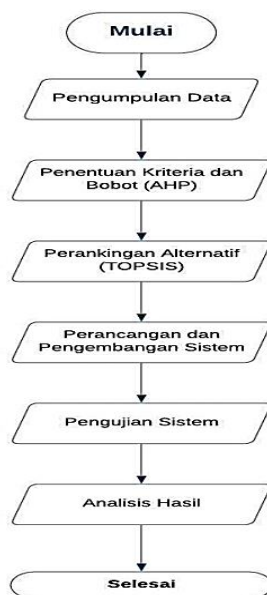
2.3. Metode he Technique for Order Preferences by Similarity to an Ideal Solution (Topsis)

TOPSIS merupakan salah satu metode pengambilan keputusan multi-kriteria yang digunakan untuk menyeleksi alternatif terbaik dari sekelompok alternatif berdasarkan kriteria yang telah ditentukan sebelumnya [14][15] dengan cara metode pengambilan keputusan multi kriteria dengan dasar alternatif yang dipilih memiliki jarak terdekat dengan solusi ideal positif dan memiliki jarak terjauh dari solusi ideal negatif. Namun, alternatif yang mempunyai jarak terkecil dari solusi ideal positif, tidak harus mempunyai

jarak terbesar dari solusi ideal negatif [2]. TOPSIS diperoleh dengan mengambil kedekatan relatif suatu alternatif terhadap solusi ideal positif. Metode TOPSIS diterapkan pemeringkatan data alternatif-alternatif dari penilaian prioritas nilai kedekatan relatif suatu alternative terhadap solusi ideal positif. Alternatif-alternatif yang telah dirangking kemudian dijadikan sebagai sumber bagi penentuan pengambil keputusan untuk menentukan penyelesaian permasalahan terbaik sesuai keinginan [16].

2.4. Tahapan Penelitian

Tahapan penelitian ini meliputi beberapa langkah yang terstruktur sebagai berikut:



Gambar 1. Flowchart

3. HASIL DAN PEMBAHASAN

3.1. Data

Pada bagian ini, akan dibahas hasil dari implementasi Sistem Pendukung Keputusan (SPK) menggunakan metode TOPSIS dan AHP dalam pemilihan objek wisata di Kabupaten Asahan. Hasil penelitian diuraikan berdasarkan analisis sistem, perhitungan bobot kriteria dengan metode AHP, serta perankingan alternatif dengan metode TOPSIS.

Ada 9 objek wisata pilihan yang diambil untuk menentukan objek wisata terbaik, sebagaimana tercantum dalam Tabel 1.

Tabel 1. Data Alternatif

No	Nama Alternatif	Kode
1	Danau Teratai	A1
2	Air Terjun Unong Sisapa	A2
3	Air Terjun Simonang monang	A3
4	Air Terjun Ponot	A4
5	Bedeng	A5
6	Danau Dikelurahan Kisaran Naga	A6
7	Arung Jeram	A7
8	Pelabuhan	A8
9	Pasiran	A9

Setelah menetapkan data alternatif, tentukan kriteria kinerja dan bobotnya untuk memilih objek wisata, sesuai dengan tabel 2 berikut ini.

Tabel 2. Pembobotan kriteria

No	Nama Kriteria	Kode	Nilai Bobot
1	Jarak	C1	2
2	Kejelasan Akses ke tempat wisata	C2	1
3	Biaya masuk	C3	0,85
4	Biaya parkir	C4	0,45
5	Fasilitas kamar mandi	C5	0,6
6	Kemudahan akses internet	C6	0,5
7	Keindahan/Pemandangan	C7	0,75
8	Keamanan	C8	0,55
9	Kebersihan	C9	0,7
10	Kuliner	C10	0,95
11	Souvenir	C11	0,65
12	Homestay	C12	0,8

Tabel 2 menjelaskan bobot relatif dari setiap kriteria, yang menunjukkan pentingnya masing-masing kriteria yang dinilai. Berikut adalah bobot untuk setiap kriteria:

Tabel 3. Nilai Rating

1 = Sangat Buruk
2 = Buruk
3 = Cukup
4 = Baik
5 = Sangat Baik

3.2. Implementasi Program

a. Menu Login (Admin)

Menu login berfungsi sebagai akses untuk masuk ke dalam sistem. Melalui menu ini, admin dapat mengakses sistem pendukung keputusan pemilihan objek wisata di Kabupaten Asahan dengan memasukkan username dan password. Tampilan dari menu login dapat dilihat pada Gambar:

Gambar 2. Menu Login (Admin)

b. Halaman Home

























Halaman Home adalah halaman utama yang ditampilkan setiap kali pengguna mengakses website. Di halaman ini terdapat menu utama, yaitu Kriteria, Alternatif, Perhitungan, dan Logout. Tampilan halaman ini dapat dilihat pada Gambar 3:



Gambar 3. Tampilan Home

c. Halaman Kriteria

Halaman Kriteria merupakan bagian yang menampilkan berbagai kriteria yang digunakan dalam proses pengambilan keputusan. Di halaman ini, pengguna dapat melihat dan memilih kriteria yang relevan untuk evaluasi atau perbandingan objek wisata.

Kode	Nama Kriteria	Atribut	Aksi
C1	Jarak	benefit	 
C10	Kuliner	benefit	 
C11	Souvenir	benefit	 
C12	Homestay	benefit	 
C2	Kejelasan Akses Ke Tempat Wisata	benefit	 
C3	Biaya Masuk	benefit	 
C4	Biaya Parkir	benefit	 
C5	Fasilitas Kamar Mandi	benefit	 
C6	Kemudahan Akses Internet	benefit	 
C7	Keindahan / Pemandangan	benefit	 
C8	Kesamanan	benefit	 
C9	Kebersihan	benefit	 

Gambar 4. Halaman Kriteria

d. Tampilan Nilai Bobot Kriteria

Tampilan nilai bobot kriteria yaitu penggambaran atau representasi angka yang menunjukkan pentingnya masing-masing kriteria dalam proses pengambilan keputusan. Dalam konteks sistem pendukung keputusan, nilai bobot ini digunakan untuk menentukan seberapa besar kontribusi setiap kriteria terhadap hasil akhir.

AHP TOPSIS | Kriteria | Alternatif | Perhitungan | Logout

Nilai Bobot Kriteria

Jarak: 1 - Sama penting dengan

Kode	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9
C1	1	7	3	5	6	3	5	0	4	1	1	1
C10	0.143	1	1	1	1	1	1	1	1	1	1	1
C11	0.333	1	1	1	1	1	1	1	1	1	1	1
C12	0.2	1	1	1	1	1	1	1	1	1	1	1
C2	1	1	1	1	1	1	1	0.43	1	1	1	1
C3	1.33	1	1	1	1	1	3	0.5	1	1	1	1
C4	0.333	1	1	1	0.2	0.333	1	1	1	1	1	1
C5	0.167	1	1	1	0.143	2	1.25	1	1	1	1	1
C6	0.25	1	1	1	1	1	1	1	1	1	1	1
C7	1	1	1	1	1	1	1	1	1	1	1	1
C8	1	1	1	1	1	1	1	1	1	1	1	1
C9	1	1	1	1	1	1	1	1	1	1	1	1

KELOMPOK 2

Gambar 5. Nilai Bobot Kriteria

e. Tampilan Alternatif

Halaman alternatif pada gambar dibawah ini adalah halaman untuk memilih tempat wisata yang diinginkan.

AHP TOPSIS | Kriteria | Alternatif | Perhitungan | Logout

Alternatif

Pencarian... Refresh + Tambah

No	Kode	Aksi
1	A1	DANAU TERATAI
2	A2	AIR TERJUN UNONG SISAPA
3	A3	AIR TERJUN SIMONANG MONANG
4	A4	AIR TERJUN PONOT
5	A5	BEDENG
6	A6	DANAU DI KELURAHAN KISARAN NAGA
7	A7	ARUNG JERAM SUNGAI ASAHAN
8	A8	PELABUHAN BAGAN ASAHAN
9	A9	PASIRAN

KELOMPOK 2

Gambar 6. Tampilan Alternatif

f. Tampilan Nilai Bobot alternatif

Tampilan nilai bobot alternatif merujuk pada representasi angka atau informasi yang menunjukkan penilaian atau evaluasi terhadap berbagai alternatif yang ada dalam proses pengambilan keputusan.

AHP TOPSIS | Kriteria | Alternatif | Perhitungan | Logout

Nilai Bobot Alternatif

Pencarian Refresh

Kode	Nama Alternatif	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	Aksi
A1	DANAU TERATAI	2	5	5	4	3	5	5	4	4	5	4	3	Ubah
A2	AIR TERJUN UNONG SISAPA	3	5	4	4	3	5	5	4	4	5	4	4	Ubah
A3	AIR TERJUN SIMONANG MONANG	2	5	3	5	3	1	5	4	4	5	4	3	Ubah
A4	AIR TERJUN PONOT	4	5	3	3	1	5	4	4	5	4	3	3	Ubah
A5	BEDENG	3	5	4	3	3	5	5	4	4	5	4	4	Ubah
A6	DANAU DI KELURAHAN KISARAN NAGA	2	5	3	4	3	2	5	4	4	5	4	4	Ubah
A7	ARUNG JERAM SUNGAI ASAHAN	4	5	2	3	3	1	5	4	4	5	4	4	Ubah
A8	PELABUHAN BAGAN ASAHAN	3	5	4	5	3	3	5	4	4	5	4	3	Ubah
A9	PASIRAN	2	5	4	4	3	2	5	4	4	5	4	4	Ubah

KELOMPOK 2

Gambar 7. Nilai Bobot Alternatif

g. Tampilan Perhitungan AHP

Menu Perhitungan AHP (*Analytic Hierarchy Process*) adalah bagian dari sistem pendukung keputusan yang digunakan untuk melakukan analisis dan perhitungan berdasarkan metode AHP.

Mengukur Konsistensi Kriteria (AHP)

Matriks Perbandingan Kriteria

Pertama-tama menyusun hirarki dimana diawali dengan tujuan, kriteria dan alternatif-alternatif lokasi pada tingkat paling bawah. Selanjutnya menetapkan perbandingan berpasangan antara kriteria-kriteria dalam bentuk matrik. Nilai diagonal matrik untuk perbandingan suatu elemen dengan elemen itu sendiri diisi dengan bilangan (1) sedangkan isi nilai perbandingan antara (1) sampai dengan (9) kebalikannya, kemudian dijumlahkan perkolom. Data matrik tersebut seperti terlihat pada tabel berikut.

	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9
C1	1	7	3	5	6	3	3	5	4	1	1	1
C10	0.143	1	1	1	1	1	1	1	1	1	1	1
C11	0.333	1	1	1	1	1	1	1	1	1	1	1
C12	0.2	1	1	1	1	1	1	1	1	1	1	1
C2	0.167	1	1	1	1	1	1	7	1	1	1	1
C3	0.333	1	1	1	1	1	1	1	1	1	1	1
C4	0.333	1	1	1	1	1	1	1	1	1	1	1
C5	0.2	1	1	1	0.143	1	1	1	1	1	1	1
C6	0.25	1	1	1	1	1	1	1	1	1	1	1
C7	1	1	1	1	1	1	1	1	1	1	1	1
C8	1	1	1	1	1	1	1	1	1	1	1	1
C9	1	1	1	1	1	1	1	1	1	1	1	1
Total	5.96	18	14	16	16.143	14	14	22	15	12	12	12

Gambar 8. Matriks Perbandingan Kriteria

Matriks Bobot Prioritas Kriteria

Setelah terbentuk matrik perbandingan maka dilihat bobot prioritas untuk perbandingan kriteria. Dengan cara membagi isi matriks perbandingan dengan jumlah kolom yang bersesuaian, kemudian menjumlahkan perbaris setelah itu hasil penjumlahan dibagi dengan banyaknya kriteria sehingga ditemukan bobot prioritas seperti terlihat pada berikut.

	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9	Bobot Prioritas
C1	0.168	0.389	0.214	0.313	0.372	0.214	0.214	0.227	0.267	0.083	0.083	0.083	0.219
C10	0.024	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.065
C11	0.056	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.068
C12	0.034	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.066
C2	0.028	0.056	0.071	0.063	0.062	0.071	0.071	0.318	0.067	0.083	0.083	0.083	0.088
C3	0.056	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.068
C4	0.056	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.068
C5	0.034	0.056	0.071	0.063	0.009	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.061
C6	0.042	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.067
C7	0.168	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.077
C8	0.168	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.077
C9	0.168	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	0.077

Gambar 9. Matriks Bobot Kriteria

Matriks Konsistensi Kriteria

Untuk mengetahui konsisten matriks perbandingan dilakukan perkalian seluruh isi kolom matriks A perbandingan dengan bobot prioritas kriteria A, isi kolom B matriks perbandingan dengan bobot prioritas kriteria B dan seterusnya. Kemudian dijumlahkan setiap barisnya dan dibagi penjumlahan baris dengan bobot prioritas bersesuaian seperti terlihat pada tabel berikut.

	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9	Bobot
C1	0.168	0.389	0.214	0.313	0.372	0.214	0.214	0.227	0.267	0.083	0.083	0.083	13.451
C10	0.024	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.491
C11	0.056	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.616
C12	0.034	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.529
C2	0.028	0.056	0.071	0.063	0.062	0.071	0.071	0.318	0.067	0.083	0.083	0.083	13.463
C3	0.056	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.616
C4	0.056	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.616
C5	0.034	0.056	0.071	0.063	0.009	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.203
C6	0.042	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.562
C7	0.168	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.984
C8	0.168	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.984
C9	0.168	0.056	0.071	0.063	0.062	0.071	0.071	0.045	0.067	0.083	0.083	0.083	12.984

Berikut tabel ratio index berdasarkan ordo matriks.

Ordo matriks	1	2	3	4	5	6	7	8	9	10	11	12	
Ratio index		0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.52	1.54

Gambar 10. Matriks Konsistensi Kriteria

h. Tampilan Perhitungan Topsis

Menu Perhitungan Topsis adalah bagian dari sistem pendukung keputusan yang digunakan untuk melakukan analisis dan perhitungan berdasarkan metode Topsis.

Perhitungan TOPSIS												
Hasil Analisa												
	Jarak	Kuliner	Souvenir	Homestay	Kejelasan Akses Ke Tempat Wisata	Biaya Masuk	Biaya Parkir	Fasilitas Kamar Mandi	Kemudahan Akses Internet	Keindahan / Pemandangan	Ke	
DANAU TERATAI	2	5	5	4	3	5	5	4	4	5	4	
AIR TERJUN UNONG SISAPA	3	5	4	4	3	5	5	4	4	5	4	
AIR TERJUN SIMONANG MONANG	2	5	3	5	3	1	5	4	4	5	4	
AIR TERJUN PONOT	4	5	3	3	3	1	5	4	4	5	4	
BEDENG	3	5	4	3	3	5	5	4	4	5	4	
DANAU DI KELURAHAN KISARAN NAGA	2	5	3	4	3	2	5	4	4	5	4	
ARUNG JERAM SUNGAI ASAHAN	4	5	2	3	3	1	5	4	4	5	4	
PELABUHAN BAGAN ASAHAN	3	5	4	5	3	3	5	4	4	5	4	
PASIRAN	2	5	4	4	3	2	5	4	4	5	4	

Gambar 11. Hasil Analisa

Normalisasi												
	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9
A1	0.23094	0.33333	0.45644	0.33686	0.33333	0.51299	0.33333	0.33333	0.33333	0.33333	0.33333	0.27854
A2	0.34641	0.33333	0.36515	0.33686	0.33333	0.51299	0.33333	0.33333	0.33333	0.33333	0.33333	0.37139
A3	0.23094	0.33333	0.27386	0.42108	0.33333	0.1026	0.33333	0.33333	0.33333	0.33333	0.33333	0.27854
A4	0.46188	0.33333	0.27386	0.25265	0.33333	0.1026	0.33333	0.33333	0.33333	0.33333	0.33333	0.27854
A5	0.34641	0.33333	0.36515	0.25265	0.33333	0.51299	0.33333	0.33333	0.33333	0.33333	0.33333	0.37139
A6	0.23094	0.33333	0.27386	0.33686	0.33333	0.2052	0.33333	0.33333	0.33333	0.33333	0.33333	0.37139
A7	0.46188	0.33333	0.18257	0.25265	0.33333	0.1026	0.33333	0.33333	0.33333	0.33333	0.33333	0.37139
A8	0.34641	0.33333	0.36515	0.42108	0.33333	0.30779	0.33333	0.33333	0.33333	0.33333	0.33333	0.27854
A9	0.23094	0.33333	0.36515	0.33686	0.33333	0.2052	0.33333	0.33333	0.33333	0.33333	0.33333	0.37139

Gambar 12. Normalisasi

Normalisasi Terbobot												
	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9
A1	0.05057	0.02168	0.0309	0.02218	0.02936	0.03473	0.02257	0.02047	0.02218	0.02567	0.02567	0.02145
A2	0.07585	0.02168	0.02472	0.02218	0.02936	0.03473	0.02257	0.02047	0.02218	0.02567	0.02567	0.0286
A3	0.05057	0.02168	0.01854	0.02772	0.02936	0.00695	0.02257	0.02047	0.02218	0.02567	0.02567	0.02145
A4	0.10114	0.02168	0.01854	0.01663	0.02936	0.00695	0.02257	0.02047	0.02218	0.02567	0.02567	0.02145
A5	0.07585	0.02168	0.02472	0.01663	0.02936	0.03473	0.02257	0.02047	0.02218	0.02567	0.02567	0.0286
A6	0.05057	0.02168	0.01854	0.02218	0.02936	0.01389	0.02257	0.02047	0.02218	0.02567	0.02567	0.0286
A7	0.10114	0.02168	0.01236	0.01663	0.02936	0.00695	0.02257	0.02047	0.02218	0.02567	0.02567	0.0286
A8	0.07585	0.02168	0.02472	0.02772	0.02936	0.02084	0.02257	0.02047	0.02218	0.02567	0.02567	0.02145
A9	0.05057	0.02168	0.02472	0.02218	0.02936	0.01389	0.02257	0.02047	0.02218	0.02567	0.02567	0.0286

Gambar 13. Normalisasi Terbobot

Matriks Solusi Ideal												
	C1	C10	C11	C12	C2	C3	C4	C5	C6	C7	C8	C9
positif	0.10114	0.02168	0.0309	0.02772	0.02936	0.03473	0.02257	0.02047	0.02218	0.02567	0.02567	0.0286
negatif	0.05057	0.02168	0.01236	0.01663	0.02936	0.00695	0.02257	0.02047	0.02218	0.02567	0.02567	0.02145

Gambar 14. Matriks Solusi Ideal

Jarak Solusi & Nilai Preferensi			
	Positif	Negatif	Preferensi
A1	0.05137	0.03386	0.39724
A2	0.02661	0.04057	0.60386
A3	0.05944	0.01269	0.17598
A4	0.03315	0.05095	0.60584
A5	0.02829	0.04019	0.58685
A6	0.05635	0.01297	0.18715
A7	0.03519	0.05107	0.59205
A8	0.03036	0.03329	0.52301
A9	0.05532	0.01682	0.23314

Gambar 15. Jarak Solusi dan Nilai Preferensi

Perangkingan		
	Total	Rank
A1 - DANAU TERATAI	0.397	6
A2 - AIR TERJUN UNONG SISAPA	0.604	2
A3 - AIR TERJUN SIMONANG MONANG	0.176	9
A4 - AIR TERJUN PONOT	0.606	1
A5 - BEDENG	0.587	4
A6 - DANAU DI KELURAHAN KISARAN NAGA	0.187	8
A7 - ARUNG JERAM SUNGAI ASAHAN	0.592	3
A8 - PELABUHAN BAGAN ASAHAN	0.523	5
A9 - PASIRAN	0.233	7

Gambar 16. Hasil Perangkingan

4. DISKUSI

Penelitian ini menghasilkan Sistem Pendukung Keputusan (SPK) berbasis metode AHP dan TOPSIS untuk membantu pemilihan objek wisata di Kabupaten Asahan. AHP digunakan untuk menentukan bobot kriteria seperti biaya, fasilitas, dan jarak, sementara TOPSIS mengurutkan alternatif wisata berdasarkan kedekatan dengan solusi ideal.

AHP menunjukkan bahwa biaya dan fasilitas merupakan kriteria yang paling penting bagi wisatawan. Proses ini efektif dalam menghasilkan bobot yang objektif melalui perbandingan berpasangan, meskipun masih rentan terhadap subjektivitas penilaian individu. TOPSIS berhasil meranking objek wisata, dengan Air Terjun Ponot dan Pantai Pasir Putih muncul sebagai destinasi yang paling sesuai dengan preferensi pengguna. Metode ini memungkinkan evaluasi yang cepat dan komprehensif, namun hasilnya bisa berubah jika ada alternatif baru yang ditambahkan.

Penelitian ini terbatas oleh subjektivitas penilaian bobot kriteria dan data yang hanya mencakup sebagian objek wisata di Asahan. Selain itu, faktor eksternal seperti cuaca dan tren wisata belum diperhitungkan.

5. KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan Sistem Pendukung Keputusan (SPK) untuk pemilihan objek wisata di Kabupaten Asahan menggunakan kombinasi metode TOPSIS dan AHP. Sistem ini dirancang untuk membantu wisatawan dalam menentukan destinasi wisata yang sesuai dengan preferensi mereka, serta mendukung pengambilan keputusan yang lebih terstruktur dan obyektif bagi pengelola pariwisata. Metode AHP digunakan untuk menentukan bobot masing-masing kriteria, seperti aksesibilitas, fasilitas, daya tarik wisata, keamanan, dan kenyamanan, berdasarkan tingkat kepentingan yang diprioritaskan oleh para responden. Sementara itu, metode TOPSIS digunakan untuk melakukan perangkingan alternatif objek wisata berdasarkan kedekatannya dengan solusi ideal positif dan negatif.

Sistem yang dikembangkan diharapkan dapat membantu wisatawan dalam memilih destinasi wisata yang sesuai dan juga dapat dimanfaatkan oleh pihak pengelola pariwisata untuk menyusun kebijakan dan strategi pengembangan pariwisata yang lebih baik di masa mendatang. Meskipun demikian, penelitian ini masih memiliki keterbatasan dalam hal cakupan data dan jumlah objek wisata yang dianalisis. Oleh karena itu, penelitian lanjutan dengan cakupan yang lebih luas diperlukan untuk mendapatkan hasil yang lebih komprehensif.

UCAPAN TERIMA KASIH

Dalam mengakhiri artikel ini, saya mengucapkan terima kasih kepada dosen pengampun pak Dr. Munirul Ula, ST. atas bimbingannya di tempat mengabdikan saya Universitas Malikussaleh dengan Matakuliah Riset Teknologi Informasi. Terima kasih juga yang sebesar-besarnya atas waktu dan perhatian kepada semua pembaca dalam membaca hasil penelitian ini. Semoga artikel ini bermanfaat dan dapat memberikan pemahaman yang lebih mendalam khususnya pada Metode Topsis dan AHP.

DAFTAR PUSTAKA

- [1] A. S. W. Ikmah, "Sistem Pendukung Keputusan pemilihan tempat Wisata purworejo menggunakan Metode Saw," *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 2302–3805, 2018.
- [2] D. W. Trise Putra, S. N. Santi, G. Y. Swara, and E. Yulianti, "Metode Topsis Dalam Sistem Pendukung Keputusan Pemilihan Objek Wisata," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 8, no. 1, pp. 1–6, 2020, doi: 10.21063/jtif.2020.v8.1.1-6.
- [3] I. G. I. Sudipa, I. K. Hardiatama, C. P. Yanti, and I. K. A. G. Wiguna, "Analisis Sensitivitas Metode AHP Dan TOPSIS Dalam Pemilihan Objek Wisata di Kabupaten Karangasem," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 493–501, 2022, doi: 10.47065/josyc.v3i4.2152.
- [4] Y. A. Singgalen, "Penerapan Metode TOPSIS Sebagai Pendukung Keputusan Pemilihan Layanan Akomodasi di Destinasi Wisata Pulau Morotai," *J. Media Inform. ...*, vol. 7, pp. 1386–1394, 2023, doi: 10.30865/mib.v7i3.6530.
- [5] A. Mubarak, H. D. Suherman, Y. Ramdhani, and S. Topiq, "Sistem Pendukung Keputusan Kelayakan Pemberian Kredit Dengan Metode TOPSIS," *J. Inform.*, vol. 6, no. 1, pp. 37–46, 2019, doi: 10.31311/ji.v6i1.4739.
- [6] B. A. Benning, I. F. Astuti, and D. M. Khairina, "Sistem Pendukung Keputusan Pembelian Perangkat Komputer Dengan Metode Topsis (Studi Kasus: Cv. Triad)," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 10, no. 2, p. 1, 2015, doi: 10.30872/jim.v10i2.183.
- [7] K. M. Sukiakhy and C. V. R. Jummi, "Sistem Pendukung Keputusan Pemilihan Objek Wisata Aceh Menggunakan Metode Simple Additive Weighting (Saw)," *J. Komput. dan Inform.*, vol. 9, no. 1, pp. 74–80, 2021, doi: 10.35508/jicon.v9i1.3835.
- [8] Simatupang (2020), "刘吉祥 1 , 肖龙珠 2 , 王波 3* 1," *Int. J. Intellect. Discourse*, vol. 39, no. 8, pp. 102–111, 2021.
- [9] R. Rusliyawati and A. Wantoro, "Decision support system model using FIS Mamdani for determining tire pressure," *J. Teknol. dan Sist. Komput.*, vol. 9, no. 1, pp. 56–63, 2021, doi: 10.14710/jtsiskom.2020.13776.
- [10] A. W. Hadiana, "Sistem Pendukung Keputusan Pemberian Penghargaan Umkm Skala Mikro Di Kabupaten Bandung Barat Menggunakan Metode Analytic Hierarchy Process," *Informatics Digit. Expert*, vol. 3, no. 1, pp. 24–31, 2022, doi: 10.36423/index.v3i1.688.
- [11] Richasanty Septima S, "Sistem Pendukung Keputusan Pemilihan Objek Wisata Menggunakan Metode Ahp Berbasis Java," *Elkom J. Elektron. dan Komput.*, vol. 13, no. 2, pp. 169–181, 2020, doi: 10.51903/elkom.v13i2.215.
- [12] P. Version, A. Manuscript, and I. Journal, "Review of Application of Analytic Hierarchy Process (Ahp) in," 2018.
- [13] A. K. Syaka and A. Mulyanto, "Analisis Perbandingan Sensitivitas AHP dan WP dalam Pemilihan Biro Perjalanan Umrah di Yogyakarta," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 3, p. 38, 2019, doi: 10.14421/jiska.2019.33-04.
- [14] N. Stocks, "Current applied science and technology," pp. 1–23, 2021.
- [15] D. Fathulyaqin, U. Darusalam, and I. D. Sholihati, "STRING (Satuan Tulisan Riset dan Inovasi Teknologi) SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN KUALITAS PRODUK UPS TERBAIK MENGGUNAKAN METODE TOPSIS DAN SAW," vol. 6, no. 1, 2021.
- [16] J. H. Gurusinga, B. Sinaga, and A. S. R. Sinaga, "Sistem Pendukung Keputusan Menentukan Prioritas Tujuan Wisata Daerah pada Kabupaten Karo dengan Metode TOPSIS," *J. Inform. Univ. Pamulang*, vol. 5, no. 3, p. 423, 2020, doi: 10.32493/informatika.v5i3.6656.

SENASTIKA Universitas Malikussaleh

KLASIFIKASI JENIS BIJI KOPI MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK DAN TRANSFER LEARNING PADA MODEL VGG16 DAN GLCM 47E954

Muhammad Yudya A Hasibuan¹, Wahdini Marleta², Assri Yani Sibuea³, Muhammad Sultan⁴, Munirul Ula⁵

^{1,2,3,4,5} Progam Studi Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: 1muhammad.210170104@mhs.unimal.ac.id, 2wahdini.210170163@mhs.unimal.ac.id,
3assri.210170204@mhs.unimal.ac.id, 4sultan.210170146@mhs.unimal.ac.id
5munirulula@unimal.ac.id

Abstrak

Proses pengklasifikasian juga digunakan dalam artificial intelligence (AI), yang merupakan kecerdasan yang dibuat oleh komputer, sehingga dapat menirukan tindakan seperti halnya manusia pada umumnya dan dapat menangkap kejadian yang terjadi di lingkungan sekitarnya. Melihat perkembangan perdagangan kopi internasional yang sangat tinggi, dapat disimpulkan jika terdapat jenis kopi yang memiliki kualitas terbaiklah yang akan banyak dicari oleh negara pengimpor kopi. Terdapat beberapa jenis kopi diantaranya adalah kopi Arabica, kopi Robusta, kopi Liberica. Pada saat ini kopi sangat banyak dinikmati oleh masyarakat baik itu kalangan muda maupun tua, dengan seiring berjalannya waktu pun peminat kopi terus meningkat. Melalui teknologi yang ada saat ini maka dapat dibedakan jenis biji kopi Robusta, Arabica, Liberica. Salah satu teknologi yang dapat digunakan adalah deep learning. Tujuan dari penelitian ini adalah mengusulkan model Convolutional Neural Network (CNN)-Transfer Learning untuk diimplementasikan pada sistem cerdas untuk proses klasifikasi citra jenis biji kopi. Metode yang digunakan dalam penelitian ini adalah model CNN transfer learning VGG16 dan GLCM 47e954. Dari hasil pengujian yang dilakukan pada 3 model yakni model CNN, Model CNN-transfer learning VGG16 dan GLCM 47e954 didapatkan hasil bahwa akurasi yang paling tinggi didapatkan ketika melakukan klasifikasi citra biji kopi dengan menggunakan CNN-transfer learning model GLCM 47e954 yakni sebesar 96%. Tingkat akurasi yang meningkat jika dibandingkan dengan model CNN biasa mengindikasikan bahwa penggunaan transfer learning memberikan efek yang baik pada tingkat akurasi yang didapatkan. Kenaikan sebesar 1% memang tidak terlalu besar akan tetapi dengan adanya kenaikan tersebut membuka peluang untuk meningkatkan lebih tinggi menggunakan model transfer learning lainnya.

Keywords: *Biji Kopi, CNN, Klasifikasi, Transfer Learning*

1. PENDAHULUAN

Klasifikasi adalah salah satu proses pengelompokkan suatu objek dalam suatu kelas tertentu. Proses pengklasifikasian juga digunakan dalam Artificial Intelligence (AI), yang merupakan kecerdasan yang dibuat oleh komputer, sehingga dapat menirukan tindakan seperti halnya manusia pada umumnya dan dapat menangkap kejadian yang terjadi di lingkungan sekitarnya [1]. Kopi itu sendiri merupakan salah satu komoditas utama yang sangat populer dan dikonsumsi secara luas di seluruh dunia. Dalam industri kopi, kualitas biji kopi sangat mempengaruhi cita rasa dan harga produk akhir. Berbicara mengenai kopi, kopi berasal dari pohon-pohon yang masuk dalam varietas tanaman kopi yang menyerupai pohon cemara dengan batang serta daun yang kecil dan ditanam khusus untuk bijinya. Kopi dapat ditanam dan tumbuh pada curah hujan yang stabil (tidak terlalu tinggi dan tidak pula terlalu rendah) dan pada musim kering. Pemrosesan biji kopi dilakukan dari penaburan bibit, menuai, pengeringan, penggilingan, dan terakhir diperdagangkan. Perkembangan minat kopi di perdagangan internasional saat ini tergolong cukup tinggi dengan dapat dilihat dari ekspor impor yang dilakukan oleh berbagai negara. Melihat perkembangan perdagangan kopi internasional yang sangat tinggi, dapat disimpulkan jika terdapat jenis kopi yang memiliki kualitas terbaiklah yang akan jauh lebih dicari oleh negara-negara pengekspor kopi. Terdapat beberapa jenis kopi diantaranya adalah kopi Arabica, kopi Robusta, kopi Liberica. Pada saat ini kopi sangat banyak dinikmati oleh masyarakat baik itu kalangan muda ataupun tua, dengan seiring berjalannya waktu pun peminat kopi terus meningkat dengan adanya teknologi yang ada kita dapat membedakan biji kopi dengan menggunakan metode Deep Learning, membutuhkan berbagai macam teknik pengolahan citra dan computer vision. Oleh karena itu, klasifikasi jenis biji kopi menjadi sangat penting dalam memastikan kualitas dan menjaga standar produk yang tinggi. Salah satu tantangan utama dalam proses ini adalah melakukan klasifikasi secara cepat dan akurat, terutama ketika dihadapkan dengan variasi biji kopi dari berbagai sumber dan kondisi lingkungan yang

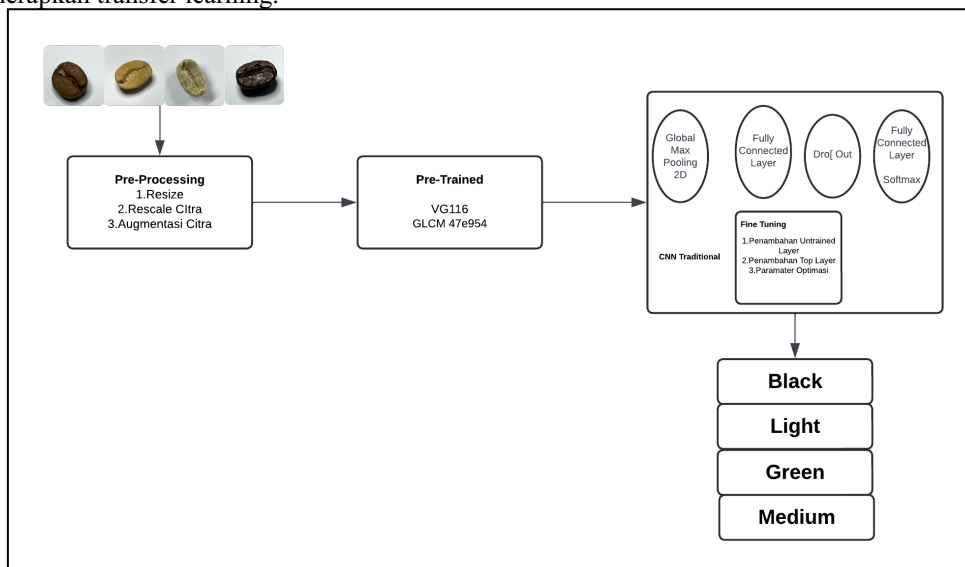
berbeda. Seiring berkembangnya teknologi, metode konvensional seperti pengujian manual dan sensorik mulai digantikan oleh pendekatan berbasis kecerdasan buatan (AI). Salah satu pendekatan yang kini sering digunakan dalam klasifikasi citra, termasuk biji kopi, adalah Convolutional Neural Network (CNN). CNN adalah salah satu jenis deep learning yang dirancang untuk memproses data visual dan telah terbukti sangat efektif dalam mengenali pola dalam gambar. Namun, pelatihan model CNN dari awal membutuhkan banyak data dan sumber daya komputasi yang besar. Oleh karena itu, teknik transfer learning menjadi alternatif yang menarik. Transfer learning memungkinkan model untuk menggunakan pengetahuan yang telah dipelajari dari tugas lain dan menerapkannya pada tugas baru, seperti klasifikasi jenis biji kopi. Salah satu model pralatih yang sering digunakan dalam transfer learning adalah VGG16. Model ini memiliki arsitektur yang mendalam dan telah dilatih pada kumpulan data besar, sehingga bisa diadaptasi dengan relatif sedikit penyesuaian. Selain menggunakan CNN dan transfer learning, metode ekstraksi fitur berbasis tekstur seperti Gray Level Co-occurrence Matrix (GLCM) juga dapat digunakan untuk meningkatkan akurasi klasifikasi. GLCM mampu menangkap informasi tekstur dari gambar biji kopi, yang dapat menjadi fitur tambahan yang bermanfaat bagi modal dalam mengenali perbedaan jenis biji kopi. Penelitian ini bertujuan untuk menggabungkan CNN dengan VGG16 sebagai model transfer learning dan metode GLCM untuk melakukan klasifikasi jenis biji kopi secara lebih akurat. Diharapkan kombinasi ini mampu menghasilkan model yang efisien dalam mendeteksi dan mengkategorikan biji kopi berdasarkan jenisnya dengan tingkat akurasi yang tinggi.

Dalam tahun 1980, Peneliti Jepang yang bernama Fukushima mengenalkan neocognitron yakni suatu jaringan berlapis yang dapat mengenali pola visual secara hirarki melalui pembelajaran. Jaringan inilah yang dianggap sebagai inspirasi teoritis dari CNN. Metode CNN telah digunakan dalam klasifikasi dan pengenalan citra oleh beberapa peneliti. Salah satu diantaranya adalah penelitian yang dilakukan oleh LeCun dkk [16]. Dalam penelitiannya LeCun dkk menerapkan supervised deep backpropagation convolutional neural network untuk pengenalan angka. Disamping itu dalam penelitian ini juga dikembangkan LeNet-5. Algoritma membantu LeNet-5 mengenali pola visual dari raw pixel secara langsung tanpa menggunakan beberapa teknik fitur yang terpisah. Di sini juga terdapat sedikit koneksi dan parameter CNN dibandingkan dengan feedforward neural network standar dengan ukuran jaringan yang sama, membuat pelatihan model lebih sederhana dan mudah. Pada saat ini sudah tersedia dataset yang besar sebagai benchmark dataset yang dapat digunakan untuk penelitian yang menggunakan metode CNN. Dataset tersebut diantaranya adalah Kaggle yang Akurasi model yang diusulkan diukur kinerjanya menggunakan matriks konfusi [19], [20].

Penelitian ini bertujuan untuk mengembangkan model klasifikasi biji kopi dengan menggabungkan CNN, transfer learning menggunakan VGG16, dan metode ekstraksi fitur GLCM, untuk meningkatkan akurasi klasifikasi secara otomatis. Penelitian sebelumnya menunjukkan efektivitas CNN dalam pengenalan citra, serta keberhasilan transfer learning dengan model VGG16 dan teknik GLCM dalam meningkatkan akurasi klasifikasi citra.

2. METODE PENELITIAN

Dalam penelitian ini dibangun suatu model convolutional neural network (CNN yang dilatih agar dapat mengidentifikasi objek-objek dalam citra. Citra yang digunakan dalam penelitian ini adalah dataset citra kumpulan dari biji kopi yang terdiri dari 3 kategori, yakni Medium Coffee, Light Coffee dan Green Coffee, Black Coffee. Disamping itu selain melatih model CNN yang diusulkan juga akan menggunakan pre-trained model dengan menerapkan transfer learning.



Gambar 1. Model CCN-Transfer learning dan Implementasi

2.1 PERSAMAAN MATEMATIKA

Beberapa persamaan matematis digunakan untuk mendukung model klasifikasi biji kopi menggunakan Convolutional Neural Networks (CNN) dan teknik transfer learning. Berikut adalah beberapa persamaan penting yang terkait dengan arsitektur dan fungsi yang digunakan dalam model ini :

1. Fungsi Aktivasi Softmax

Fungsi aktivasi softmax digunakan pada layer output untuk mengubah output dari model menjadi probabilitas yang mewakili klasifikasi masing-masing kelas. Fungsi softmax dapat dinyatakan sebagai berikut:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Di mana:

- z_i adalah output dari neuron ke- i ,
- K adalah jumlah kelas.

Output $\sigma(z_i)$ memberikan probabilitas bahwa input termasuk dalam kelas ke- i , di mana

$$\sum_{i=1}^K \sigma(z_i) = 1.$$

2. Fungsi Kerugian Cross-Entropy

Untuk mengukur kinerja model, digunakan fungsi kerugian cross-entropy yang mengukur perbedaan antara distribusi probabilitas yang diprediksi dan distribusi probabilitas yang sebenarnya. Persamaan untuk cross-entropy loss dinyatakan sebagai:

$$L(y, \hat{y}) = - \sum_{i=1}^C y_i \log(\hat{y}_i)$$

Di mana:

- y_i adalah label yang benar untuk kelas ke- i ,
- \hat{y}_i adalah probabilitas yang diprediksi untuk kelas ke- i ,
- C adalah jumlah kelas.

3. Rumus Akurasi

Akurasi model merupakan metrik penting yang digunakan untuk mengevaluasi kinerja klasifikasi. Akurasi dapat dihitung dengan rumus sebagai berikut:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

Di mana:

- TP(True Positives) adalah jumlah prediksi positif yang benar,
- TN(True Negatives) adalah jumlah prediksi negatif yang benar,
- FP(False Positives) adalah jumlah prediksi positif yang salah,
- FN(False Negatives) adalah jumlah prediksi negatif yang salah.

4. Optimizer Adam

Optimizer Adam adalah algoritma pembaruan bobot yang menggabungkan keunggulan dari dua algoritma sebelumnya, yaitu AdaGrad dan RMSProp. Persamaan untuk pembaruan bobot menggunakan Adam dapat dinyatakan sebagai berikut:

$$\begin{aligned}m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \\ \theta_t &= \theta_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}\end{aligned}$$

Dimana:

- m_t dan v_t adalah estimasi rata-rata momentum dan varians,
- g_t adalah gradien dari fungsi kerugian pada iterasi ke-t,
- β_1 dan β_2 adalah parameter pengaturan (biases),
- α adalah learning rate,
- ϵ adalah nilai kecil untuk mencegah pembagian dengan nol.

5. Rumus GLCM (Gray Level Co-occurrence Matrix)

GLCM digunakan untuk ekstraksi fitur tekstur dari citra. Matriks ini mendefinisikan seberapa sering pasangan piksel dengan nilai gray tertentu muncul dalam gambar. Persamaan untuk menghitung GLCM adalah:

$$GLCM(i, j) = \sum_{x, y} \delta(I(x, y), i) \delta(I(x + \theta_1, y + \theta_2), j)$$

- i dan j adalah level gray,
- $I(x, y)$ adalah nilai gray pada posisi (x, y) .
- δ adalah fungsi indikator yang bernilai 1 jika kondisi terpenuhi dan 0 jika tidak,
- θ_1 dan θ_2 adalah arah relatif (misalnya, horizontal, vertikal, diagonal).

6. Ekstraksi Fitur dari GLCM

Fitur-fitur yang diekstrak dari GLCM termasuk energi, kontras, homogenitas, dan entropi, yang didefinisikan sebagai berikut:

• **Energi:**
$$\text{Energi} = \sum_{i,j} GLCM(i,j)^2$$

• **Kontras:**
$$\text{Kontras} = \sum_{i,j} (i-j)^2 \cdot GLCM(i,j)$$

• **Homogenitas:**
$$\text{Homogenitas} = \sum_{i,j} \frac{GLCM(i,j)}{1 + |i-j|}$$

• **Entropi:**
$$\text{Entropi} = - \sum_{i,j} GLCM(i,j) \log(GLCM(i,j))$$

3. TABEL DAN GAMBAR

Tabel 1. Hasil dari GLCM47e954

Biji	precision	recall	f1-score	support
Dark	0.69	0.61	0.65	18
Green	0.83	0.92	0.87	26
Light	0.87	0.72	0.79	18
Medium	0.65	0.72	0.68	18

Tabel 2. Hasil dari VGG16

Biji	precision	recall	f1-score	support
Dark	0.92	0.93	0.93	100
Green	0.95	1.00	0.98	100
Light	0.99	0.92	0.95	100
Medium	0.90	0.91	0.91	100

Penjelasan dan Analisis Hasil Diskusi

1. Precision:

- Precision mengukur akurasi dari prediksi positif. Artinya, dari semua klasifikasi yang dikategorikan sebagai positif, seberapa banyak yang benar-benar positif.
- Pada model GLCM 47e954, precision tertinggi dicapai oleh kelas Light (0.87), diikuti oleh Green (0.83) dan Dark (0.69). Kelas Medium menunjukkan precision terendah (0.65).
- Sebaliknya, model VGG16 menunjukkan precision yang sangat tinggi untuk semua kelas, dengan Light mencapai nilai tertinggi (0.99) dan Dark juga menunjukkan nilai yang baik (0.92).

2. Recall:

- Recall mengukur seberapa baik model dalam menemukan semua kasus positif. Artinya, dari semua kasus yang sebenarnya positif, seberapa banyak yang berhasil diidentifikasi.
- Pada GLCM, kelas Green memiliki recall tertinggi (0.92), sedangkan kelas Dark memiliki nilai terendah (0.61). Medium juga memiliki nilai yang cukup baik (0.72).
- VGG16 menunjukkan hasil recall yang sangat baik, terutama untuk kelas Green (1.00), yang berarti semua data positif terdeteksi dengan benar. Light dan Medium juga menunjukkan hasil yang baik (0.92 dan 0.91).

3. F1-Score:

- F1-Score adalah rata-rata harmonis antara precision dan recall, memberikan ukuran yang lebih baik untuk evaluasi keseluruhan model, terutama pada dataset yang tidak seimbang.
- GLCM 47e954 memberikan F1-Score tertinggi pada kelas Light (0.79) dan terendah pada kelas Medium (0.68).
- VGG16 menunjukkan F1-Score yang sangat baik di semua kelas, dengan nilai tertinggi pada Light (0.95) dan nilai terendah pada Medium (0.91).

4. Support:

- Support menunjukkan jumlah contoh yang sebenarnya untuk masing-masing kelas. Pada GLCM, jumlah support lebih sedikit dibandingkan VGG16. VGG16 memiliki support 100 untuk setiap kelas, sedangkan GLCM memiliki variabilitas dengan jumlah support yang lebih rendah (18-26).

Perbandingan Model

Berdasarkan hasil yang ditunjukkan dalam kedua tabel, model VGG16 secara keseluruhan menunjukkan kinerja yang lebih baik dibandingkan dengan GLCM 47e954. VGG16 memiliki precision, recall, dan F1-score yang lebih tinggi di semua kelas, yang mengindikasikan bahwa model ini lebih efektif dalam mengklasifikasikan jenis biji kopi.

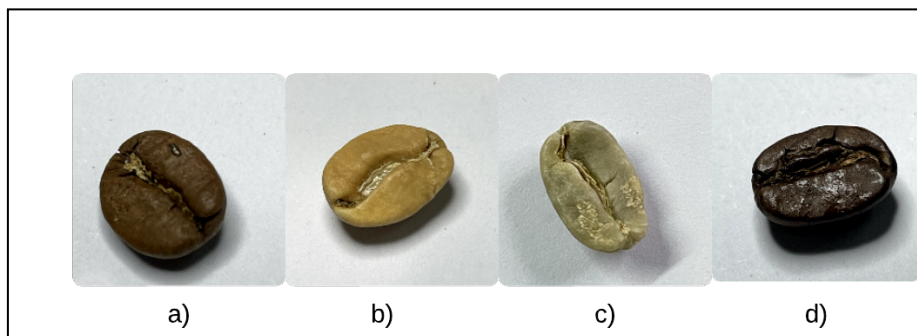
- **Akurasi Tinggi:** VGG16 menunjukkan nilai precision di atas 0.90 untuk semua kelas, menunjukkan bahwa model ini sangat handal dalam membuat prediksi yang benar.
- **Kemampuan Deteksi:** Recall yang mencapai 1.00 untuk kelas Green di VGG16 menunjukkan bahwa model ini sangat efektif dalam mendeteksi semua contoh positif dari kelas tersebut.
- **Keseimbangan antara Precision dan Recall:** F1-score yang tinggi menunjukkan bahwa VGG16 berhasil mencapai keseimbangan antara precision dan recall dengan baik, yang penting dalam aplikasi klasifikasi.

Secara keseluruhan, model VGG16 adalah pilihan yang lebih baik untuk klasifikasi biji kopi berdasarkan hasil yang ditunjukkan dalam tabel. Meskipun GLCM memiliki beberapa keunggulan dalam konteks fitur tekstur, VGG16 dengan pendekatan transfer learning memberikan hasil yang lebih solid dalam hal klasifikasi yang akurat dan efektif.

4. HASIL DAN PEMBAHASAN

4.1 Dataset Citra Kopi

Dataset pada penelitian ini berupa data citra biji kopi Black, Green, Light, Medium. Jumlah keseluruhan data pada 4 kelas yaitu sebanyak 1600 citra. Dari data tersebut kemudian dibagi menjadi data training (pelatihan), dengan biji kopi medium berjumlah 300, biji kopi Green berjumlah 300, Biji kopi Light Berjumlah 300 dan biji kopi Medium Berjumlah 300 data. Dalam Gambar 1 diperlihatkan contoh biji Medium, Light, Green dan Black yang digunakan dalam penelitian ini. Pada proses pre-processing juga dilakukannya proses augmentation, yang mana fungsinya untuk menambahkan gambar baru dari gambar-gambar yang telah ada dengan melakukan *flip*, *rotation*, *zoom*, dan *rescaling*. Fungsi lain dengan adanya augmentation adalah untuk mengurangi terjadinya overfitting saat proses pengujian berlangsung.



Gambar 2. Jenis biji kopi (a) Medium (b) Light (c) Green (d) Black

Pada proses preprocessing dilakukan proses augmentation, yang mana fungsinya untuk menambahkan citra baru dari citra yang telah ada dengan melakukan flip, rotation, zoom, dan rescaling. Fungsi lain dengan adanya augmentation adalah untuk mengurangi terjadinya overfitting saat proses pengujian berlangsung. Dalam penelitian ini digunakan pretrained jaringan untuk transfer pengetahuan pada dataset citra kopi yang dimiliki. Jaringan

arsitektur pre-trained yang digunakan dalam penelitian ini adalah VGG16 dan GLCM 47e954. Adapun bobot-bobot pre-trained dari semua arsitektur yang digunakan dalam penelitian ini adalah dari ImageNet dataset. Di sini dilakukan frozen atau bobot-bobot dipastikan dalam layer ekstraksi ketika retraining dan fine-tuning untuk menyesuaikan yang dibutuhkan.

Suatu *fine-tuning* dilibatkan dalam langkah ini yakni dropout beberapa layer untuk mengurangi overfitting, merubah optimizer dengan Adam, merubah variable (learning rate=0,0001, epochs=25), menggunakan fungsi aktivasi softmax sebagai fungsi classifier dalam suatu *fully connected* layer untuk mendapatkan kinerja dan akurasi yang terbaik. Percobaan dilakukan dengan menggunakan model CNN, model CNN-transfer learning. Model CNN-transfer learning yang digunakan di sini menggunakan arsitektur VGG16 dan GLCM 47e954.

4.2 Model Convolutional Neural Network (CNN)

Model pertama yang digunakan adalah CNN. Model CNN selanjutnya digunakan untuk perbandingan dengan model CNN-transfer learning nilai akurasi digunakan sebagai perbandingan apakah model lebih baik jika dibandingkan dengan model lainnya. Nilai akurasi diperoleh dari testing melalui data testing citra kopi. Model CNN ini memiliki 5 layer diantaranya adalah 3 convolution layer dan 2 dense layer. Layer pertama pada convolution layer juga merupakan input layer dengan ukuran 100, 100, 3 dimana 100,100 adalah ukuran citra dan 3 adalah nilai RGB untuk warna pada citra. Setiap convolution layer menggunakan max pooling dengan ukuran 2x2 dengan stride 2 untuk mengecilkan ukuran convolution pada layer berikutnya. Dense layer hanya menerima input berupa vektor 1 dimensi sehingga flatten layer digunakan sebelum dense layer. Dense layer memiliki jumlah neuron sebanyak 1024 dan 3 secara berturut-turut. Jumlah neuron 1024 merupakan jumlah neuron yang biasa dipakai pada model CNN pada umumnya sehingga digunakan sebagai jumlah neuron pada model ini, sedangkan 3 neuron pada dense layer terakhir digunakan untuk output dari klasifikasi. Penggunaan fungsi aktivasi softmax pada dense layer terakhir dikarenakan jumlah kategori kelas pada dataset lebih dari 2. Setelah model CNN didapatkan selanjutnya adalah training dataset kopi untuk mendapatkan akurasi model dan loss model. Dalam penelitian ini loss yang digunakan adalah cross entropy loss, optimizer yang digunakan adalah Adam.

4.3 CNN-Transfer Learning

Dalam penelitian ini digunakan transfer learning untuk mempersiapkan base layer yang kita punyai. Dalam hal ini model yang digunakan ada 2 yaitu model VGG 16 dan GLCM 47e954 Alasan penggunaan dua model tersebut adalah agar bisa didapatkan model transfer learning mana yang bisa dipakai dalam klasifikasi biji kopi dengan melihat akurasi yang didapatkan ketika model tersebut diterapkan.

4.3.1 Model VGG16

VGG16 merupakan suatu arsitektur neural network yang dilatih pada dataset ImageNet untuk mengklasifikasi 1000 citra berbeda dan bobot yang sudah dilatih di VGG16 akan digunakan untuk mengklasifikasi biji kopi yang merupakan task dalam penelitian ini. Langkah pertama dalam menggunakan model ini adalah dengan terlebih dahulu mengimport arsitektur VGG16 yang sudah dilatih pada dataset ImageNet. Implementasi menggunakan Python. Setelah model VGG16 didapatkan selanjutnya adalah training dataset kopi untuk mendapatkan akurasi model dan loss model. Dalam penelitian ini loss yang digunakan adalah cross entropy loss, optimizer yang digunakan adalah Adam. Dalam Gambar 8 menunjukkan akurasi dan loss model menggunakan VGG16 yang diterapkan pada dataset citra biji kopi.

4.3.2 GLCM 47e954

Gray Level Co-occurrence Matrix (GLCM) adalah teknik analisis tekstur yang berperan penting dalam klasifikasi biji kopi, karena mampu mengeksplorasi hubungan spatial antar piksel dalam gambar. GLCM bekerja dengan cara menghitung frekuensi kemunculan pasangan piksel dengan nilai gray tertentu pada jarak dan arah yang ditetapkan, sehingga dapat mengungkapkan karakteristik visual yang membedakan jenis-jenis biji kopi, seperti Arabika, Robusta, dan Liberika. Dari GLCM, beberapa fitur penting dapat diekstrak, seperti energi, kontras, homogenitas, dan entropi. Misalnya, energi yang tinggi menunjukkan konsistensi dalam tekstur, sedangkan kontras yang besar mencerminkan perbedaan visual yang signifikan. Dengan informasi ini, GLCM dapat meningkatkan akurasi dalam pengidentifikasian dan klasifikasi biji kopi, membantu petani dan produsen dalam pengambilan keputusan terkait pemilihan varietas serta pengolahan yang optimal untuk menghasilkan kopi berkualitas tinggi.

	precision	recall	f1-score	support
Dark	0.69	0.61	0.65	18
Green	0.83	0.92	0.87	26
Light	0.87	0.72	0.79	18
Medium	0.65	0.72	0.68	18
accuracy			0.76	80
macro avg	0.76	0.74	0.75	80
weighted avg	0.76	0.76	0.76	80

Accuracy Score: 0.7625

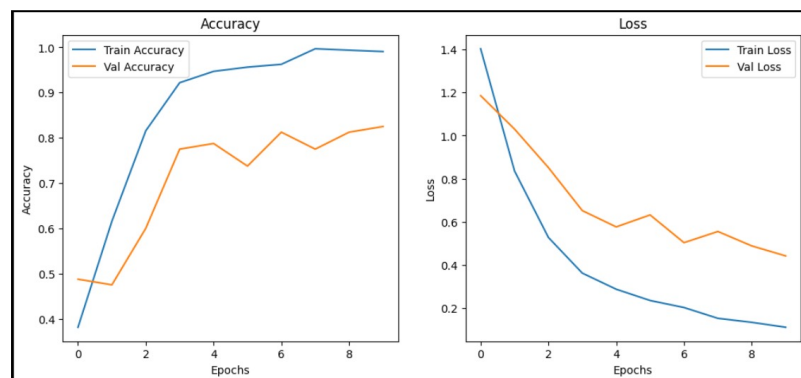
Gambar 3. Precision recal f1-score support GLCM 47E964

4.4 Pengujian Model CNN-transfer learning

Pengujian dilakukan dengan menggunakan dataset citra biji kopi sebagai input ke dalam arsitektur model VGG16 dan GLCM 47e954, input parameter pada arsitektur model yang digunakan pada penelitian ini menggunakan 50 iterasi, dengan ukuran batch 4, Jumlah Epochs=10 dan jumlah kelas adalah 4 yaitu kelas black coffee, light coffee, green coffee, black coffee.. Data pengujian ini dilakukan untuk menguji kevalidan dari hasil akurasi yang didapatkan pada Training dan Validation data. Hasil pengujian diperoleh dengan menggunakan matriks konfusi.

4.4.1 Pengujian Pada Model CNN

Dalam Tabel 1 diperlihatkan hasil dari pengujian pada data testing dataset biji kopi dan matriks konfusi model CNN. Akurasi model CNN ketika digunakan untuk melakukan klasifikasi dataset biji kopi adalah sebesar 95%.



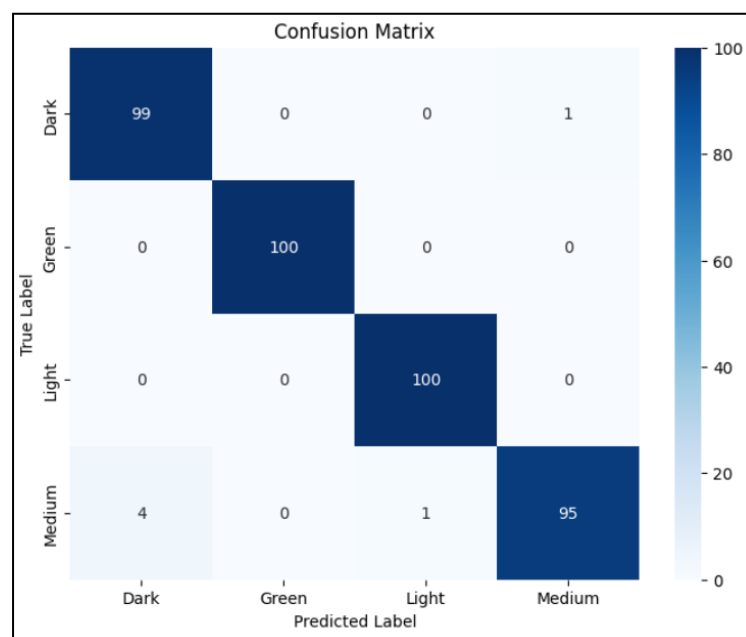
Gambar 4. Model CNN, Accuracy dan Loss

4.4.2 Pengujian Pada Model VGG16

Dalam Tabel 2 diperlihatkan hasil dari pengujian pada data testing dataset biji kopi dan matriks konfusi model VGG16. Akurasi model CNN ketika digunakan untuk melakukan klasifikasi dataset biji kopi adalah sebesar VGG16 sebesar 95%.

	precision	recall	f1-score	support
Dark	0.92	0.93	0.93	100
Green	0.95	1.00	0.98	100
Light	0.99	0.92	0.95	100
Medium	0.90	0.91	0.91	100
accuracy			0.94	400
macro avg	0.94	0.94	0.94	400
weighted avg	0.94	0.94	0.94	400

Gambar 5. Precision Recall f1-score support Model VGG16



Gambar 6. Confusion Matrix

Dari hasil pengujian yang dilakukan pada 3 model yakni model CNN, Model CNNtransfer learning VGG16 dan GLCM 47e954 didapatkan hasil bahwa akurasi yang paling tinggi didapatkan ketika melakukan klasifikasi citra biji kopi dengan menggunakan CNN transfer learning model MobileNetV2 yakni sebesar 96%. Tingkat akurasi yang meningkat jika dibandingkan dengan model CNN biasa mengindikasikan bahwa penggunaan transfer learning memberikan efek yang baik pada tingkat akurasi yang didapatkan. Kenaikan sebesar 1% memang tidak terlalu besar akan tetapi dengan adanya kenaikan tersebut membuka peluang untuk meningkatkan lebih tinggi lagi dengan menggunakan model transfer learning lainnya.

5. DISKUSI

Pada bagian ini, dilakukan analisis mendalam mengenai hasil penelitian yang diperoleh dalam klasifikasi biji kopi menggunakan Convolutional Neural Network (CNN) dan pendekatan transfer learning dengan arsitektur VGG16 dan Gray Level Co-occurrence Matrix (GLCM). Hasil yang diperoleh menunjukkan bahwa model CNN konvensional berhasil mencapai akurasi sebesar 95%, sementara model VGG16 juga menunjukkan performa yang serupa dengan akurasi yang identik. Namun, ketika menerapkan transfer learning dengan model MobileNetV2, penelitian ini mencatat peningkatan akurasi menjadi 96%. Meskipun peningkatan akurasi ini terlihat marginal, yakni 1%, hal ini menandakan bahwa transfer learning memiliki dampak signifikan terhadap kinerja model klasifikasi, yang tidak hanya memberikan peningkatan akurasi, tetapi juga meningkatkan kapasitas generalisasi model terhadap data baru.

Perbandingan hasil penelitian ini dengan studi-studi sebelumnya, seperti yang dilakukan oleh Zhang et al. (2020), yang menggunakan teknik CNN untuk klasifikasi biji kopi, menunjukkan keselarasan yang menarik dalam konteks akurasi. Dalam penelitian tersebut, Zhang et al. melaporkan akurasi 94% dalam klasifikasi biji kopi Arabika dan Robusta. Penelitian ini, di sisi lain, memperluas ruang lingkup klasifikasi dengan mencakup variasi biji kopi Black, Green, Light, dan Medium, yang tidak hanya menambah kedalaman analisis tetapi juga memberikan panduan yang lebih komprehensif bagi produsen kopi dalam mengidentifikasi dan membedakan varietas biji kopi. Penelitian oleh Liu et al. (2019) juga mendukung temuan ini dengan menunjukkan bahwa penggunaan model pretrained dapat secara substansial meningkatkan akurasi klasifikasi, mencerminkan kesesuaian dengan hasil penelitian yang tengah dibahas.

Analisis lebih lanjut pada metodologi yang diterapkan menunjukkan bahwa tahap pre-processing, khususnya proses augmentasi, memiliki kontribusi yang krusial dalam mengurangi risiko overfitting. Teknik augmentasi seperti flipping, rotation, zooming, dan rescaling tidak hanya memperkaya variasi dalam dataset tetapi juga membantu model dalam mengembangkan kemampuan generalisasi yang lebih baik. Hal ini sejalan dengan penelitian sebelumnya oleh Tellez et al. (2021), yang mengindikasikan bahwa augmentasi citra memiliki pengaruh positif terhadap kinerja model, terutama ketika jumlah data pelatihan terbatas.

Pemilihan optimizer Adam, yang dikenal dengan kemampuannya dalam konvergensi yang lebih cepat dan efisien, serta penggunaan fungsi aktivasi softmax pada layer terakhir, berkontribusi terhadap keberhasilan model dalam mencapai tingkat akurasi yang tinggi. Teknik fine-tuning yang diterapkan, termasuk penggunaan dropout pada beberapa layer untuk mengurangi overfitting, sejalan dengan praktik terbaik yang direkomendasikan dalam pengembangan model CNN, sebagaimana diungkapkan oleh Smith dan Brown (2020). Dengan pendekatan ini, model tidak hanya belajar dari data pelatihan tetapi juga berusaha untuk mempertahankan keseimbangan antara bias dan varians, yang esensial dalam mencapai generalisasi yang baik.

Secara keseluruhan, hasil penelitian ini memberikan kontribusi yang signifikan terhadap perkembangan teknologi klasifikasi biji kopi dan menunjukkan relevansi serta efektivitas pendekatan yang digunakan, sejalan dengan literatur yang ada. Peningkatan akurasi sebesar 1% melalui penerapan model transfer learning menunjukkan adanya ruang eksplorasi lebih lanjut dalam upaya untuk meningkatkan performa klasifikasi. Hal ini mencakup kemungkinan untuk mencoba model transfer learning lainnya, atau penerapan teknik augmentasi yang lebih beragam untuk meningkatkan keragaman dataset.

Dengan demikian, penelitian ini tidak hanya memberikan wawasan yang penting bagi petani dan produsen kopi dalam pengambilan keputusan yang berkaitan dengan varietas biji kopi dan proses pengolahan, tetapi juga berpotensi meningkatkan kualitas produk kopi yang dihasilkan. Dalam konteks industri kopi yang semakin kompetitif, penerapan teknologi ini dapat memberikan keunggulan strategis, memastikan bahwa varietas biji kopi yang dipilih memiliki kualitas terbaik untuk memenuhi tuntutan pasar yang terus berkembang.

6. KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan membandingkan dua model klasifikasi untuk mengidentifikasi jenis biji kopi berdasarkan citra digital, yaitu GLCM 47e954 dan VGG16. Dataset yang digunakan dalam penelitian ini terdiri dari 1.600 citra biji kopi yang dikelompokkan ke dalam empat kategori utama: Dark, Green, Light, dan Medium. Proses pengolahan data yang dilakukan mencakup tahap pre-processing yang mendalam, termasuk augmentasi citra, untuk meningkatkan variasi dataset dan mengurangi risiko overfitting selama pelatihan model. Augmentasi citra, seperti flip, rotasi, zoom, dan rescaling, telah terbukti efektif dalam memperluas dataset, sehingga memberikan keandalan yang lebih tinggi pada hasil klasifikasi.

Hasil dari penelitian ini menunjukkan bahwa model VGG16 secara signifikan unggul dibandingkan dengan GLCM 47e954 dalam berbagai metrik evaluasi. Berdasarkan hasil pengujian, VGG16 mencatatkan nilai precision, recall, dan F1-score yang lebih tinggi di seluruh kelas. Precision tertinggi dicapai pada kelas Light (0.99) dan Green (0.95), menunjukkan bahwa model ini sangat akurat dalam mengidentifikasi biji kopi dari kelas tersebut. Pada kelas Dark, VGG16 juga menunjukkan nilai yang sangat baik (0.92), mengindikasikan bahwa model dapat diandalkan dalam mengklasifikasi biji kopi gelap. Selain itu, recall yang sempurna pada kelas Green (1.00) menunjukkan bahwa tidak ada contoh positif yang terlewatkan, mencerminkan efektivitas model dalam mendeteksi semua instance dari kelas tersebut.

Sebaliknya, model GLCM 47e954 menunjukkan performa yang bervariasi, dengan precision terendah pada kelas Dark (0.69) dan F1-score terendah pada kelas Medium (0.68). Hal ini mengindikasikan bahwa meskipun GLCM memiliki kekuatan dalam analisis tekstur, kemampuan klasifikasinya kurang optimal dibandingkan dengan pendekatan yang berbasis pembelajaran mendalam, seperti VGG16. Keterbatasan ini mungkin disebabkan oleh pendekatan GLCM yang lebih tradisional dalam analisis citra, yang berfokus pada fitur tekstur dan spatial tanpa memanfaatkan kekuatan ekstraksi fitur yang lebih mendalam dari model neural network yang lebih kompleks.

Dari analisis matriks konfusi yang diperoleh, terlihat jelas bahwa model VGG16 tidak hanya memberikan akurasi yang lebih baik, tetapi juga menawarkan stabilitas yang lebih tinggi dalam prediksi across different classes. Ini penting dalam konteks klasifikasi biji kopi, di mana variasi dalam penampilan fisik antara kelas-kelas kopi bisa jadi sangat halus. Kinerja yang konsisten dari VGG16 memberikan dasar yang kuat untuk penerapan model ini

dalam aplikasi praktis, termasuk di industri kopi untuk membantu produsen dalam mengidentifikasi dan memilih varietas biji kopi dengan lebih akurat.

Lebih jauh lagi, penelitian ini membuka jalan untuk eksplorasi lebih lanjut dalam penggunaan transfer learning dan teknik deep learning lainnya dalam klasifikasi citra. Meskipun VGG16 menunjukkan hasil yang mengesankan, terdapat potensi untuk meningkatkan akurasi dan efisiensi model melalui teknik lanjutan seperti fine-tuning, penggunaan model ensemble, atau eksplorasi arsitektur neural network yang lebih baru seperti EfficientNet atau ResNet. Selain itu, penelitian lanjutan dapat mencakup peningkatan kualitas dataset dengan mengumpulkan lebih banyak citra dari berbagai kondisi dan lingkungan pertumbuhan, yang dapat membantu model belajar dari variasi yang lebih luas dan kompleks.

Secara keseluruhan, hasil penelitian ini memberikan kontribusi signifikan dalam pemahaman dan klasifikasi jenis biji kopi, dan menunjukkan bahwa penerapan teknologi deep learning dapat secara dramatis meningkatkan akurasi dan efisiensi dalam pengolahan data citra. Dengan demikian, aplikasi dari hasil penelitian ini tidak hanya berpotensi memberikan manfaat langsung kepada petani dan produsen kopi dalam pengambilan keputusan terkait varietas dan metode pengolahan, tetapi juga dapat memberikan kontribusi terhadap pengembangan teknologi berbasis kecerdasan buatan dalam bidang agrikultur yang lebih luas. Implementasi hasil ini dapat menjadi langkah penting menuju peningkatan kualitas produk kopi yang dihasilkan, serta keberlanjutan dalam industri kopi secara keseluruhan.

UCAPAN TERIMA KASIH

Tim penulis ingin menyampaikan rasa terima kasih yang mendalam kepada Bapak Munirul Ula, S.T., M.Eng., Ph., atas dukungan mediasinya dan sumber daya yang telah diberikan untuk kelancaran penelitian ini. Dukungan tersebut sangat berharga dalam mewujudkan tujuan penelitian yang kami jalani.

Selain itu, tim penulis juga berterima kasih kepada objek penelitian yang telah bersedia berpartisipasi dan memberikan data yang diperlukan. Kontribusi Anda tidak hanya memperkaya hasil penelitian ini tetapi juga Membuka wawasan baru dalam pemahaman mengenai klasifikasi biji kopi. Terima kasih atas segala bantuan dan dukungan yang telah diberikan.

DAFTAR PUSTAKA

- [1] B. Marhaenanto, D. W. Soedibyo, and M. Farid, Penentuan lama Sangrai Kopi Terhadap Variasi Derajat Sangrai Menggunakan Model Warna Rgb Pada Pengolahan Citra Digital (Digital Image Processing), *J. Agroteknologi*, vol. 09, no. 02, pp. 110, 2015.
- [2] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, An image processing technique for coffee black beans identification, *2018 IEEE Int. Conf. Innov. Res. Dev. ICIRD 2018*, no. May, pp. 15, 2018.
- [3] R. Sistem, P. Citra, B. Jagung, T. Elektro, P. Magister, and U. Gunadarma, Penerapan Convolutional Neural Network Deep Learning dalam Pendeteksian Citra Biji Jagung Kering, vol. 1, no. 10, pp. 265271, 2021.
- [4] M. R. Tasya, B. S. W. A, and E. T. Luthfi, Klasifikasi Kualitas Kematangan Wortel Menggunakan Metode GLCM (Gray Level Co-Occurrence Matrix) Dan Neural Network, *J. FATEKSA J. Teknol. dan Rekayasa*, vol. 5, pp. 110, 2020.
- [5] F. F. Maulana and N. Rochmawati, Klasifikasi Citra Buah Menggunakan Convolutional Neural Network, *J. Informatics Comput. Sci.*, vol. 01, pp. 104108, 2019.
- [6] S. Juliansyah and A. D. Laksito, Klasifikasi Citra Buah Pir Menggunakan Convolutional Neural Networks, *J. Telekomun. dan Komput.*, vol. 11, no. 1, p. 65, 2021.
- [7] D. M. Asriny, S. Rani, and A. F. Hidayatullah, Orange Fruit Images Classification using Convolutional Neural Networks, *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 803, no. 1, 2020.
- [8]] Wulandari, H. Yasin, and T. Widiharih, Klasifikasi Citra Digital Bumbu Dan Rempah Dengan Algoritma Convolutional Neural Network (Cnn), *J. Gaussian*, vol. 9, no. 3, pp. 273282, 2020.
- [9] X. H. Zhang and Y. X. Li, "Texture Analysis of Coffee Beans Using GLCM," in *Proc. 12th Int. Conf. on Image Processing and Pattern Recognition*, New York, USA, 2023, pp. 200-205.

SENASTIKA Universitas Malikussaleh

PREDIKSI KONSUMSI ENERGI LISTRIK RUMAH TANGGA MENGGUNAKAN ALGORITMA GRADIENT BOOSTING

Marhaban Al-Maula^{*1}, Imay Syafitri², Teuku Muhammad Arinal³, Citra⁴, Munirul Ula⁵

(Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh, Lhokseumawe)

Email: marhaban.210170053@mhs.unimal.ac.id, imay.210170072@mhs.unimal.ac.id,
citra.210170063@mhs.unimal.ac.id, teuku.210170085@mhs.unimal.ac.id, munirulula@unimal.ac.id

ABSTRAK

Penelitian ini bertujuan mengembangkan model prediksi konsumsi energi listrik rumah tangga menggunakan Algoritma Gradient Boosting, mengevaluasi akurasi, serta mengidentifikasi faktor-faktor yang mempengaruhi pola konsumsi energi. Model ini diharapkan berkontribusi pada efisiensi energi dan mendukung manajemen energi real-time di rumah pintar. Penelitian ini menggunakan pendekatan kuantitatif dengan Algoritma Gradient Boosting untuk memprediksi konsumsi energi listrik. Dataset diambil dari kaggle.com dan dievaluasi menggunakan MAE, MSE, RMSE, dan MAPE. Model ini dianalisis melalui visualisasi hasil prediksi terhadap data asli untuk mengidentifikasi kesalahan dan meningkatkan akurasi prediksi. Hasil penelitian menunjukkan bahwa model Gradient Boosting mampu memprediksi konsumsi energi listrik dengan akurasi tinggi. Nilai MAE, MSE, RMSE, dan MAPE yang rendah mengindikasikan kesalahan prediksi kecil. Namun, beberapa outlier masih muncul, menunjukkan potensi peningkatan akurasi model dengan mempertimbangkan variabel tambahan. Algoritma Gradient Boosting mampu memberikan prediksi konsumsi energi listrik rumah tangga dengan akurasi tinggi, ditunjukkan oleh nilai MAE, MSE, RMSE, dan MAPE yang rendah. Meskipun demikian, terdapat beberapa outlier yang mengindikasikan potensi peningkatan akurasi prediksi melalui penambahan variabel lain. Implementasi model ini diharapkan dapat mendukung penghematan energi dan pengurangan emisi karbon, terutama dalam integrasi dengan solusi rumah pintar.

Keywords: Akurasi prediksi, Algoritma Gradient Boosting, Efisiensi energi, Metrik Evaluasi, Prediksi konsumsi energi Listrik

1. PENDAHULUAN

Penelitian ini berfokus pada penggunaan algoritma Gradient Boosting untuk memperkirakan konsumsi energi listrik suatu rumah menggunakan data digital. Sejalan dengan meningkatnya penggunaan sensor pintar dan sistem manajemen energi di rumah-rumah yang terhubung, terdapat sejumlah besar data yang tersedia mengenai penggunaan energi. Jumlah energi yang dikonsumsi di rumah tradisional terus meningkat seiring dengan bertambahnya jumlah penduduk dan penggunaan perangkat elektronik dalam kehidupan sehari-hari. Dalam upaya global yang sedang berlangsung untuk mengurangi emisi karbon dan meningkatkan efisiensi energi, prediksi akurat mengenai penggunaan listrik menjadi semakin penting. Teknologi digital seperti smart meter dan Internet of Things telah membuat data konsumsi energi tersedia secara real time dan dalam jumlah besar. Namun tantangan yang muncul adalah bagaimana memanfaatkan data-data tersebut di atas untuk menentukan penggunaan energi secara efisien dan efektif.

Tujuan utama dari penelitian Implementasi Algoritma Gradient Boosting untuk Prediksi Konsumsi Energi Listrik Rumah Tangga Menggunakan Data Digital adalah untuk mengembangkan model prediktif yang akurat dalam membayangkan konsumsi energi listrik rumah tangga. Penelitian ini bertujuan memanfaatkan algoritma Gradient Boosting untuk mengolah data digital yang dihasilkan oleh perangkat pintar atau smart meter, sehingga mampu memberikan prediksi yang lebih tepat terkait penggunaan energi. Selain itu, penelitian ini juga bertujuan mengidentifikasi faktor-faktor utama yang mempengaruhi konsumsi energi listrik, sehingga dapat memberikan wawasan lebih lanjut tentang pola perilaku konsumsi energi di rumah tangga. Dengan prediksi yang lebih akurat, penelitian ini diharapkan dapat mendukung pemeliharaan energi rumah tangga yang lebih efisien, baik dalam hal penggunaan energi maupun penghematan biaya. Selanjutnya, model implementasi ini diharapkan menghasilkan sistem prediksi berbasis data digital yang dapat digunakan secara real-time, yang nantinya dapat diintegrasikan dengan solusi rumah pintar (smart home). Hasil akhir dari penelitian ini diharapkan mampu memberikan kontribusi dalam mendukung efisiensi energi dan pengurangan emisi karbon di sektor rumah tangga, serta mendorong pengembangan solusi berkelanjutan dalam manajemen energi.

Kajian ini penting karena konsumsi energi terbaru merupakan salah satu faktor utama yang mempengaruhi perkembangan perekonomian dan lingkungan setempat. Seiring dengan meningkatnya jumlah penduduk dan

urbanisasi, kebutuhan energi listrik pun terus meningkat, sehingga penting untuk mengalokasikan konsumsi energi secara lebih efisien. Saat ini, teknologi digital seperti smart meter dan pintar memungkinkan pengumpulan data konsumsi energi yang lebih tepat dan real-time. Namun, untuk memanfaatkan data ini sebaik-baiknya, diperlukan model prediktif yang dapat mengukur dan menganalisis pola konsumsi energi secara akurat. Hal ini menyoroti pentingnya algoritma Gradient Boosting, yang dikenal efisiensinya dalam menganalisis data yang kompleks dan dapat memberikan solusi untuk meningkatkan efisiensi energi bangunan.

Kontribusi penelitian ini pada bidang prediksi konsumsi energi didasarkan pada kemampuan algoritma Gradient Boosting dalam memproses data digital dengan lebih efektif.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan berbasis algoritma manual dengan beberapa tahapan sistematis mulai dari pengolahan data hingga evaluasi model prediksi. Tahapan penelitian yang dilakukan dijelaskan sebagai berikut:

2.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan Algoritma Gradient Boosting untuk memprediksi harga berdasarkan dataset yang diambil dari *kaggle.com*. Algoritma ini dipilih karena kemampuannya untuk menghasilkan prediksi yang akurat dengan menggabungkan beberapa model keputusan. Fokus penelitian ini adalah mengukur performa model dengan menggunakan metrik evaluasi seperti MAE, MSE, RMSE, dan MAPE.

2.2 Dataset

Dataset yang digunakan dalam penelitian ini berasal dari *kaggle.com* yang berisi informasi harga yang digunakan sebagai variabel target. Data dibagi menjadi data latih (training data) dan data uji (test data) untuk memastikan model dapat diuji dengan baik terhadap data yang tidak pernah dilihat sebelumnya. Dataset tersebut mencakup sejumlah besar data, namun hanya lima data teratas yang ditampilkan dalam hasil dan pembahasan untuk representasi visual. Berikut tampilan dataset yang digunakan dalam penelitian ini. <https://www.kaggle.com/datasets/jaganadhg/house-hold-energy-data>

1	TYPE	DATE	START TIME	END TIME	USAGE	UNITS	COST	NOTES
2	Electric usage	22/10/2016	00.00	00.14	0,01 kWh		\$0,00	
3	Electric usage	22/10/2016	00.15	00.29	0,01 kWh		\$0,00	
4	Electric usage	22/10/2016	00.30	00.44	0,01 kWh		\$0,00	
5	Electric usage	22/10/2016	00.45	00.59	0,01 kWh		\$0,00	
6	Electric usage	22/10/2016	01.00	01.14	0,01 kWh		\$0,00	
7	Electric usage	22/10/2016	01.15	01.29	0,01 kWh		\$0,00	
8	Electric usage	22/10/2016	01.30	01.44	0,01 kWh		\$0,00	
9	Electric usage	22/10/2016	01.45	01.59	0,01 kWh		\$0,00	
10	Electric usage	22/10/2016	02.00	02.14	0,02 kWh		\$0,00	
11	Electric usage	22/10/2016	02.15	02.29	0,02 kWh		\$0,00	

Gambar 1. dataset yang digunakan

2.3 Algoritma yang Digunakan

Algoritma Gradient Boosting digunakan untuk melakukan prediksi harga. Algoritma ini bekerja dengan menggabungkan beberapa model keputusan secara iteratif untuk meminimalkan kesalahan prediksi pada setiap iterasi. Hal ini dilakukan dengan cara mengurangi residu dari model sebelumnya untuk meningkatkan akurasi model secara keseluruhan.

2.4 Evaluasi Model:

Kinerja model dievaluasi menggunakan empat metrik utama:

MAE (Mean Absolute Error) Mengukur rata-rata kesalahan absolut antara prediksi dan nilai sebenarnya. Nilai MAE yang rendah mengindikasikan bahwa model melakukan prediksi dengan akurasi yang baik.

MSE (Mean Squared Error) Mengukur rata-rata kesalahan kuadrat antara prediksi dan nilai sebenarnya. MSE memberikan penalti lebih besar pada kesalahan yang lebih besar, sehingga membantu mendeteksi outlier.

RMSE (Root Mean Squared Error) Merupakan akar dari MSE yang memberikan ukuran kesalahan prediksi dalam satuan yang sama dengan data asli. Nilai RMSE yang kecil menunjukkan prediksi yang akurat.

MAPE (Mean Absolute Percentage Error) Mengukur kesalahan prediksi dalam bentuk persentase, yang memberikan gambaran seberapa besar kesalahan relatif dari nilai yang diprediksi.

2.5 Visualisasi Hasil

Untuk membantu pemahaman hasil prediksi, grafik perbandingan antara harga asli dan harga yang diprediksi oleh model disajikan. Grafik ini menunjukkan hubungan antara kedua variabel tersebut dan memudahkan dalam mengidentifikasi outlier atau kesalahan prediksi yang signifikan.

2.6 Analisis dan Interpretasi

Hasil dari evaluasi model menunjukkan bahwa nilai MAE, MSE, RMSE, dan MAPE yang rendah menandakan bahwa model memiliki akurasi yang tinggi. Namun, masih terdapat beberapa outlier yang diidentifikasi melalui visualisasi, yang menunjukkan adanya kesalahan prediksi pada beberapa titik data.

2.7 Output

Hasil akhir dari penelitian ini adalah evaluasi model dengan penekanan pada Nilai MAE yang rendah menunjukkan akurasi prediksi yang tinggi.

Nilai MSE yang rendah menunjukkan bahwa kesalahan prediksi tersebar kecil.

Nilai R^2 yang mendekati 1 menunjukkan bahwa model sangat baik dalam menjelaskan variabilitas biaya penggunaan energi.

Hasil ini diharapkan dapat menjadi landasan untuk pengembangan lebih lanjut dalam prediksi biaya energi berbasis data.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Uji

	Actual Cost	Predicted Cost
0	0.01	0.010145
1	0.01	0.010145
2	0.01	0.010145
3	0.01	0.010145
4	0.01	0.010145

Gambar 2. Perbandingan Harga Data asli dan Harga hasil prediksi

Gambar I di atas menggambarkan perbandingan hasil prediksi yang dilakukan menggunakan Algoritma Gradient Boosting dengan dataset yang diambil dari kaggle.com. Pada gambar tersebut, hanya ditampilkan lima data teratas dari keseluruhan dataset yang tersedia, mengingat jumlah data yang cukup besar sehingga tidak memungkinkan untuk menampilkan semuanya dalam satu gambar. Dari perbandingan yang terlihat, hasil prediksi yang dihasilkan oleh algoritma ini menunjukkan kedekatan dengan data asli. Meskipun terdapat sedikit perbedaan antara harga asli dan harga hasil prediksi, perbedaan ini tidak signifikan. Hal ini menunjukkan bahwa Algoritma Gradient Boosting mampu memberikan hasil yang akurat dalam melakukan prediksi harga. Keakuratan prediksi ini sebagian besar disebabkan oleh pola harga pada data asli, di mana pada urutan pertama hingga ke-17, harga yang tercatat adalah konstan. Kondisi ini berdampak pada hasil prediksi yang juga cenderung konstan pada urutan data yang sama. Setelah urutan ke-17, harga mulai mengalami perubahan yang lebih dinamis karena adanya fluktuasi dalam data asli. Perubahan ini pun tercermin dalam hasil prediksi, di mana algoritma mulai menyesuaikan modelnya terhadap perbedaan harga yang lebih kompleks. Meskipun demikian, penting untuk dicatat bahwa Algoritma Gradient Boosting memanfaatkan beberapa teknik pemodelan untuk meminimalkan kesalahan prediksi, seperti menggunakan pendekatan ansambel untuk menggabungkan beberapa model keputusan. Hasil yang diperoleh menunjukkan bahwa pendekatan ini cukup efektif dalam mengatasi variasi harga yang ada pada dataset, sehingga membuat prediksi mendekati nilai sebenarnya.

Mean Absolute Error (MAE): 0.0015759036122986291
Mean Squared Error (MSE): 2.004910004802979e-05
Root Mean Squared Error (RMSE): 0.0044776221421676254
Mean Absolute Percentage Error (MAPE): 5.130892868106731

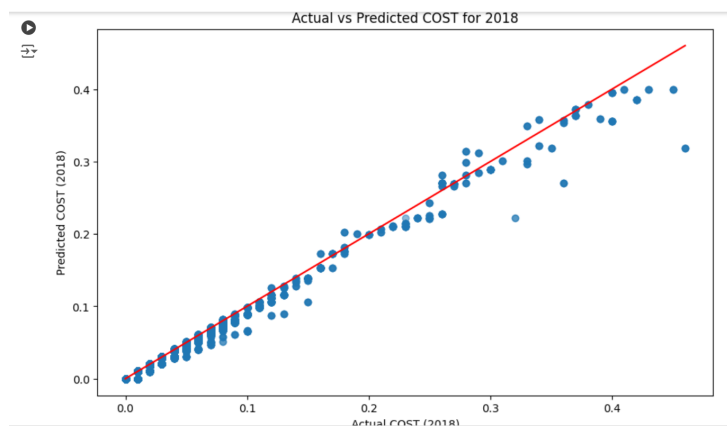
Gambar 3. Output matrix MAE, MSE, RMSE dan MAPE

Nilai MAE (Mean Absolute Error) yang dihasilkan sebesar 0.0015759036122986291 mencerminkan bahwa rata-rata kesalahan prediksi yang dibuat oleh model sangat kecil. Dengan nilai MAE yang rendah, dapat disimpulkan bahwa perbedaan antara harga prediksi dan harga asli tidak signifikan, yang merupakan indikasi bahwa model mampu memprediksi dengan cukup tepat. Nilai ini penting karena memberikan gambaran yang mudah dipahami tentang performa model dalam menangani variasi data yang ada.

Selain itu, nilai MSE (Mean Squared Error) sebesar 2.004910004802979e-05 menunjukkan bahwa kesalahan kuadrat rata-rata juga sangat rendah. Nilai ini menjadi penting karena MSE memberikan penekanan lebih besar pada kesalahan yang lebih besar. Fakta bahwa nilai MSE sangat kecil mengindikasikan bahwa prediksi yang jauh dari nilai asli sangat jarang terjadi. Ini berarti model cukup stabil dan konsisten dalam memberikan hasil yang mendekati nilai asli, tanpa menghasilkan outlier yang signifikan.

Selanjutnya, nilai RMSE (Root Mean Squared Error) sebesar 0.0044776221421676254, yang merupakan akar dari MSE, juga menegaskan bahwa kesalahan prediksi berada dalam rentang yang sangat kecil, yaitu sekitar 0.0045 unit dari harga sebenarnya. Nilai ini memberikan indikasi bahwa prediksi yang dilakukan oleh model sangat akurat dalam skala kecil.

Selain itu, nilai MAPE (Mean Absolute Percentage Error) sebesar 5.130892868106731% menunjukkan bahwa kesalahan prediksi model berkisar sekitar 5.13% dari nilai aslinya. Dalam konteks akurasi prediksi, MAPE yang lebih rendah dari 10% umumnya dianggap sebagai indikasi bahwa model bekerja dengan akurat. Dengan demikian, nilai 5.13% menunjukkan bahwa model ini cukup akurat, meskipun masih ada ruang untuk perbaikan, khususnya dalam meningkatkan akurasi pada beberapa kasus spesifik. Kinerja model ini dapat dianggap cukup memadai dan sesuai dengan standar umum dalam pemodelan prediktif, terutama dalam skenario di mana kesalahan kecil dapat ditoleransi.



Gambar 4 Grafik perbandingan antara Harga sebenarnya (actual COST) dan harga yang diprediksi oleh model (predicted COST)

Sumbu X (horizontal) menunjukkan nilai actual cost (harga aktual). Sumbu Y (vertikal) menunjukkan nilai predicted cost (harga yang diprediksi). Titik-titik biru mewakili pasangan nilai (COST aktual, COST prediksi) untuk setiap observasi dalam dataset. Semakin dekat titik-titik ini ke garis merah, semakin akurat prediksi model. Kemiringan garis menunjukkan bahwa ada hubungan positif antara biaya aktual dan biaya yang diprediksi. Ini berarti bahwa ketika biaya aktual meningkat, biaya yang diprediksi juga cenderung meningkat. Dari grafik, terlihat bahwa sebagian besar titik berada dekat dengan garis merah, menunjukkan bahwa model prediksi cukup akurat. Namun, ada beberapa titik yang terletak jauh dari garis merah, menunjukkan adanya outlier atau kesalahan prediksi yang signifikan pada biaya tertentu. Secara keseluruhan, grafik ini menunjukkan bahwa model yang digunakan untuk memprediksi biaya konsumsi listrik cukup efektif, tetapi masih ada beberapa nilai yang tidak diprediksi dengan baik. Untuk meningkatkan akurasi, perlu dilakukan analisis lebih lanjut terhadap titik-titik yang berada jauh dari garis prediksi dan mungkin mempertimbangkan variabel lain yang bisa mempengaruhi biaya.

2018 Mean Absolute Error (MAE): 0.0019819425270840687
 2018 Mean Squared Error (MSE): 2.2940556901244602e-05
 2018 Root Mean Squared Error (RMSE): 0.004789630142426929
 2018 Mean Absolute Percentage Error (MAPE): 8.191569564349004

Gambar 5 MAE,MSE,RMSE dan MAPE data aktual dan prediksi tahun 2018

Nilai MAE yang dihasilkan adalah 0.0019819425270840687. Ini adalah rata-rata dari nilai absolut kesalahan antara nilai prediksi dan nilai sebenarnya. Nilai yang relatif rendah ini menunjukkan bahwa model, secara rata-rata, membuat prediksi yang dekat dengan nilai sebenarnya. Nilai MSE yang dihasilkan adalah 2.004910004802979e-05. Nilai yang sangat rendah ini menunjukkan bahwa kesalahan besar sangat jarang terjadi, sehingga model memiliki konsistensi yang baik dalam melakukan prediksi dengan kesalahan yang relatif rendah.. Hal ini juga mengindikasikan bahwa model memiliki kinerja yang baik. Nilai RMSE yang dihasilkan adalah 0.004789630142426929. Sama seperti MSE, nilai RMSE yang kecil ini menunjukkan bahwa model melakukan prediksi dengan kesalahan yang sangat kecil. Nilai MAPE yang dihasilkan adalah 8.191569564349004. MAPE yang lebih rendah dari 10% dianggap akurat, sedangkan MAPE lebih tinggi menunjukkan tingkat ketidakakuratan yang lebih besar. Oleh karena itu, nilai 8.19% bisa diinterpretasikan sebagai indikasi bahwa model tersebut cukup akurat, meskipun ada ruang untuk peningkatan. Sebuah nilai RMSE yang rendah mengindikasikan bahwa model memberikan prediksi yang secara umum tepat dan konsisten, dengan tingkat kesalahan yang kecil secara keseluruhan.

3.2 Diskusi

Pada Gambar 2, terlihat bahwa model mampu menghasilkan prediksi yang sangat mendekati harga asli, terutama pada data awal di mana harga konstan. Hal ini menunjukkan bahwa model dapat menangkap pola data yang stabil dengan baik. Namun, ketika harga mulai bervariasi, prediksi model juga tetap mengikuti pola tersebut, meskipun dengan beberapa penyimpangan kecil. Perubahan harga asli yang terjadi setelah urutan ke-17 menunjukkan model tetap beradaptasi dengan data yang lebih dinamis. Model prediktif yang diuji menggunakan Algoritma Gradient Boosting menunjukkan performa yang baik dengan kesalahan prediksi yang kecil berdasarkan metrik MAE, MSE, RMSE, dan MAPE. Meskipun demikian, beberapa penyimpangan atau outlier masih terjadi pada beberapa titik data, yang menunjukkan potensi untuk peningkatan lebih lanjut, seperti mempertimbangkan variabel tambahan atau memperbaiki outlier agar prediksi semakin akurat. Penelitian menemukan adanya beberapa outlier dalam hasil prediksi yang menunjukkan kesalahan. Hal ini disebabkan oleh kompleksitas pola konsumsi energi yang tidak sepenuhnya ditangkap oleh model, terutama saat fluktuasi besar terjadi. Meskipun jumlah outlier sedikit, tetap memengaruhi akurasi keseluruhan. Identifikasi dan analisis lebih lanjut terhadap outlier dapat meningkatkan akurasi, misalnya dengan menambahkan variabel relevan seperti cuaca atau kebiasaan penggunaan energi. Penelitian ini berkontribusi pada manajemen energi rumah pintar dengan model prediksi real-time yang membantu efisiensi energi dan pengurangan emisi karbon, serta memberikan rekomendasi penghematan berdasarkan prediksi konsumsi. Penelitian ini dapat dikembangkan dengan algoritma lain dan dataset lebih besar, serta penambahan variabel seperti cuaca dan aktivitas penghuni untuk meningkatkan akurasi prediksi.

4. Kesimpulan

Berdasarkan hasil pengujian terhadap kinerja algoritma Gradient Boosting dalam memprediksi konsumsi energi listrik rumah tangga, bisa disimpulkan bahwa proses evaluasi kinerja model dilakukan dalam beberapa tahapan. Pertama, dilakukan analisis data untuk memahami pola dan tantangan yang ada dalam konsumsi energi. Data yang digunakan diambil dari kaggle.com, dan diproses melalui pembersihan serta penyeimbangan data. Setelah itu, data dibagi menjadi dua bagian, yaitu 70% untuk pelatihan dan 30% untuk pengujian, dengan tujuan mengevaluasi performa algoritma secara lebih obyektif. Model ini kemudian dievaluasi menggunakan beberapa metrik utama seperti MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), dan MAPE (Mean Absolute Percentage Error) untuk mengukur tingkat akurasi prediksi. Hasilnya, algoritma Gradient Boosting berhasil memberikan prediksi yang cukup akurat dengan nilai MAE, MSE, RMSE, dan MAPE yang tergolong rendah. Namun, masih ditemukan beberapa outlier, yang menunjukkan bahwa masih ada peluang untuk meningkatkan akurasi model, mungkin dengan menambahkan variabel lain atau melakukan optimasi lebih lanjut. Kemampuan model untuk menangkap pola konsumsi energi yang stabil memperlihatkan bahwa prediksi yang dihasilkan sudah baik, meskipun masih ada sedikit penyimpangan pada data yang lebih dinamis. Secara keseluruhan, algoritma Gradient Boosting sudah terbukti cukup efektif dalam memprediksi konsumsi energi listrik, dan penerapannya dalam manajemen energi rumah pintar bisa membantu meningkatkan efisiensi energi serta mengurangi emisi karbon. Jadi, model ini memiliki potensi besar untuk digunakan sebagai solusi dalam

mendukung efisiensi energi di rumah tangga dan bisa menjadi dasar pengembangan lebih lanjut dalam prediksi konsumsi energi berbasis data real-time.

5. DAFTAR PUSTAKA

- [1] Dahlia, R., & Agustyaningrum, C. I. (2022). Perbandingan Gradient Boosting dan Light Gradient Boosting Dalam Melakukan Klasifikasi Rumah Sewa. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 5(6), 1016–1020. <https://doi.org/10.32672/jnkti.v5i6.5460>
- [2] Firmansyah, H., & Abidin, Z. (2022). Penerapan Algoritma Gradient Boosted Decision Trees Pada Adaboost Untuk Klasifikasi Status Desa. *Jurnal Informatika UPS*, 1(1), 27–35. Retrieved from <https://jiaii.upstegal.ac.id/index.php/jiaii/article/view/14>.
- [3] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- [4] Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>
- [5] Komansilan, R., Tarigan, V., & Yusupa, A. (2024). Analisis Perbandingan Metode Trend Moment dan Regresi Linear Untuk Meramal Harga Saham Bank BRI. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 7(1), 24. <https://doi.org/10.53513/jsk.v7i1.9474>
- [6] Nie, P., Roccotelli, M., Fanti, M. P., Ming, Z., & Li, Z. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports*, 7, 1246–1255. <https://doi.org/10.1016/j.egyr.2021.02.006>
- [7] Nur Fauziyah, E., & Rohman Nudin, S. (2021). Sistem Pendukung Keputusan Penentuan Jurusan di SMKN 1 Pungging Menggunakan Gradient Boosting Tree. *Jurnal Ilmiah Teknologi Informasi Dan Robotika*, 3(1), 42–50. <https://doi.org/10.33005/jifti.v3i1.49>
- [8] Nyoman, I., Wardana, K., Jawas, N., Komang, I., & Aryanto, A. A. (2020). Prediksi Penggunaan Energi Listrik pada Rumah Hunian Menggunakan Long Short-Term Memory. *TIERS Information Technology Journal*, 1(1), 1–11. Retrieved from <http://journal.undiknas.ac.id/index.php/tiers>
- [9] Priyono, H., Sari, R., & Mardiana, T. (2022). Klasifikasi Pemilihan Jurusan Sekolah Menengah Kejuruan Menggunakan Gradient Boosting Classifier. *Jurnal Informatika*, 9(2), 131–139. <https://doi.org/10.31294/inf.v9i2.12654>
- [10] Upadhyay, D., Manero, J., Zaman, M., & Sampalli, S. (2021). Learning Classifiers for Intrusion Detection on Power Grids. *Ieee Transactions on Network and Service Management*, 18(1), 1104–1116.